Contents lists available at ScienceDirect

Journal of Economic Behavior & Organization

journal homepage: www.elsevier.com/locate/jebo

Learning in a black box☆

Heinrich H. Nax[a,*], Maxwell N. Burton-Chellew[b], Stuart A. West[c], H. Peyton Young[d]

[a] Department of Humanities, Political and Social Sciences, ETH Zürich, Switzerland
[b] Magdalen College, Oxford & Department of Zoology, University of Oxford, The Tinbergen Building, S Parks Rd, Oxford OX1 3PS, United Kingdom
[c] Department of Zoology, University of Oxford, The Tinbergen Building, S Parks Rd, Oxford OX1 3PS, United Kingdom
[d] Department of Economics, University of Oxford, Nuffield College, New Rd, Oxford OX1 1NF, United Kingdom

ARTICLE INFO

Article history:
Received 27 July 2015
Received in revised form 4 December 2015
Accepted 8 April 2016
Available online 13 April 2016

Keywords:
Learning
Information
Public goods game

ABSTRACT

We study behavior in repeated interactions when agents have no information about the structure of the underlying game and they cannot observe other agents' actions or payoffs. Theory shows that even when players have no such information, there are simple payoff-based learning rules that lead to Nash equilibrium in many types of games. A key feature of these rules is that subjects search differently depending on whether their payoffs increase, stay constant or decrease. This paper analyzes learning behavior in a laboratory setting and finds strong confirmation for these asymmetric search behaviors in the context of voluntary contribution games. By varying the amount of information we show that these behaviors are also present even when subjects have full information about the game.

1. Introduction

Mainstream research in experimental game theory has focused primarily on situations where the structure of the game is known and the actions of other players are observable. In practice, however, many strategic environments are so large and complex that individuals are mostly in the dark about others' actions and how they affect their own payoffs. How do players learn in such situations?

To address this question, we conducted a series of experiments in which subjects played a voluntary contribution game, but they had no information about the nature of the game, and they could not observe the actions or payoffs of the other players.[1] In effect, subjects were inside a 'black box' and the only data available to them were the payoffs they received from taking different actions. We tested the key features of learning models that have recently been proposed for such settings, and compared subjects' behaviors in the black box environment with other treatments where players were given increasing

☆ Nax's research was supported by the Office of Naval Research Grant N00014-09-1-0751 and by the European Commission through the European Research Council Advanced Investigator Grant 'Momentum' 324247. Burton-Chellew's research was supported by the European Research Council. West's research was supported by the European Research Council. Young's research was supported by the Office of Naval Research Grant N00014-09-1-0751 and the Air Force Office of Scientific Research Grant FA9550-09-1-053.
* Corresponding author at: ETH Zürich, Clausiusstrasse 37-C3, 8092 Zurich, Switzerland.
E-mail address: hnax@ethz.ch (H.H. Nax).
[1] There is a related literature on low-information gambling tasks (Nevo and Erev, 2012; Erev and Haruvy, 2013; Ben Zion et al., 2010). Prior low-information experiments in games provide subjects with *some* information allowing communication (Colman et al., 2010), or regarding the structure of the game (Rapoport et al., 2002; Weber, 2003) or other players' actions (Friedman et al., 2015).

http://dx.doi.org/10.1016/j.jebo.2016.04.006
0167-2681/

amounts of information about the nature of the game and about the actions and/or payoffs of the other players. Although this information permitted more complex adaptive responses, we found that the key features of black-box learning models are still present even when subjects have full information.

There is a growing body of theoretical work that shows how equilibrium (or near-equilibrium) play can evolve from simple, adaptive rules that require no information about what other players are doing or ex ante knowledge of the game. Such rules are said to be *payoff-based* or *completely uncoupled*, because they depend solely on the pattern of an individual's received payoffs, and not on the behaviors or payoffs of anyone else.[2] Classical reinforcement models form one class of examples,[3] but in recent years many other payoff-based rules have been identified that differ from classical reinforcement in certain key respects. These rules go under a variety of names including 'win-stay-lose-shift' (Robbins, 1952; Nowak et al., 1993), 'near-far search' (Thuijsman et al., 1995; Motro and Shmida, 1995), 'exploration-exploitation' (Eiben and Schippers, 1998; Bowling and Veloso, 2002), 'trial-and-error' (Young, 2009; Marden et al., 2009), and 'probe-and-adjust' (Skyrms, 2010).

Although these rules are broadly similar to reinforcement learning models, they differ by positing important *asymmetries* in the way that subjects respond to different patterns of experienced payoffs. These rules are also quite different from belief-based learning models, which typically assume that subjects know the structure of the game and can observe other players' actions.

An essential feature of these rules is that the pattern of received payoffs triggers different modes of search: "experimentation" (present when payoffs are steady or increasing), "exploration" (triggered when payoffs decline) and "inertia" (staying with one's previous choice).[4] In this paper, we formalize and test five specific behaviors that have been suggested in the literature on payoff-based learning:

(i) *Asymmetric inertia.* Inertia is greater when payoffs are steady or increasing than when they are decreasing.[5]
(ii) *Asymmetric volatility.* Adjustments have higher variance under exploration (when payoffs decline) than under experimentation (when payoffs are steady or increasing).
(iii) *Asymmetric breadth.* Adjustments are broader under exploration than under experimentation.[6]
(iv) *Reversion.* Under both experimentation and exploration, reversion to one's previous strategy is more likely when payoffs decrease than when they are steady or increasing.
(v) *Directional bias.* When the strategy space can be linearly ordered, the direction of change tends to be reinforced: a direction that leads to higher payoffs tends to be followed by a further change in the same direction next period; a direction that leads to lower payoffs tends to be followed by a change of direction next period.

Various forms of directional bias have been discussed in the experimental literature; see among others Selten and Stoecker (1986), Selten (1998), Harstad and Selten (2013), Bayer et al. (2013), Burton-Chellew et al. (28), and Nax and Perc (2015). The other four features have been proposed in the following models, some of which are theoretical and some empirical.

1 *Near-far search* (Thuijsman et al., 1995; Motro and Shmida, 1995). This is an empirical model of bee foraging behaviour. Bees search for nectar close to their current location (within a given patch of flowers) as long as their payoffs are steady or increasing. Once a drop in the nectar stream occurs they fly far away to a new randomly chosen patch. Thus negative payoffs trigger wide-area search, whereas steady or positive payoffs lead to local search (*asymmetric breadth, asymmetric volatility*).
2 *Win-stay-lose-shift* (Robbins, 1952; Nowak et al., 1993). Subjects keep their current strategy if it has steady or increasing payoffs; otherwise they choose a new strategy at random (*asymmetric inertia*).
3 *Probe-and-adjust* (Skyrms, 2010; Huttegger and Skyrms, 2012; Huttegger et al., 2014). Subjects occasionally experiment with different strategies, which they retain if the resulting payoff is higher; otherwise they revert to their previous strategy (*reversion*).
4 *Trial-and-error* (Young, 2009). Subjects retain their current strategy with high probability if payoffs are steady or increasing. Occasionally they may try a different strategy which is retained if the resulting payoff is higher; otherwise they revert to the previous strategy. When subjects experience a decrease in payoff from their current strategy, they choose a new one at random (*asymmetric inertia, reversion*).

[2] For details about *uncoupled* and *completely uncoupled* learning rules respectively, see Hart and Mas-Colell (2003), Hart and Mas-Colell (2006), Foster and Young (2006), and Young (2009).
[3] Classical reinforcement models include Thorndike (1898), Bush and Mosteller (1953), Suppes and Atkinson (1959), Herrnstein (1970), Harley (1981), Cross (1983), Roth and Erev (1995), Erev and Rapoport (1998), Nevo and Erev (2012), and Erev and Haruvy (2013).
[4] These asymmetries are crucial for their convergence properties (Foster and Young, 2006; Germano and Lugosi, 2007; Young, 2009; Marden et al., 2009, 2014; Pradelski and Young, 2012; Skyrms, 2012; Huttegger and Skyrms, 2012; Huttegger et al., 2014).
[5] A separate question is whether inertia is affected by the *size* as well as the direction of recent payoff changes ('surprise triggers change') as proposed by Nevo and Erev (2012). We discuss this issue in more detail in Section 3.
[6] Note that adjustments can be broad (i.e. far from one's current strategy) even though the variance of these adjustments is small; thus (ii) and (iii) are distinct.

We shall show that all five features mentioned earlier are confirmed at high levels of statistical significance in the black box treatment. Moreover, even in the treatments where more information is available, none of the hypotheses can be rejected. We do not claim that these are the only relevant features of payoff-based learning, nor do we attempt to fit a parametric model to the data, as is common in much of the experimental games literature. Rather, our aim is to identify a set of qualitative search behaviors that can be examined in many other settings.

The rest of this paper contains details of the experimental set-up, the description of the learning model, and the summary of methods and results. An appendix containing experimental instructions, regression tables and figures follows the main text.

2. The experimental set-up

2.1. Data

A total of 236 subjects, in 16 separate sessions involving 12 or 16 subjects making 80 decisions each, participated in our experiments, yielding a total of 18,880 observations. In this section, we shall describe the structure of each repeated game experiment, and the different information treatments. The data underlying this study were first reported in Burton-Chellew and West (2013), which showed that fundamentally different information settings may lead to very similar patterns of play at the aggregate level. The purpose of the present paper is to identify specific adjustment patterns at the individual level that drive the macro-dynamics.

2.2. Voluntary contribution game

The voluntary contribution game is the standard game in experimental economics to study the provision of public goods (Isaac et al., 1985; Isaac and Walker, 1988; Ledyard, 1995; Chaudhuri, 2011). It proceeds as follows. Each player i in population $N = \{1, \ldots, n\}$ makes a nonnegative, real-numbered contribution, c_i, from a finite budget $B > 0$ (here $B = 40$). The vector of contributions is denoted by $\mathbf{c} = \{c_1, c_2, \ldots, c_n\}$. Given a *rate of return*, $e \geq 1$, the public good is provided in the amount $R(\mathbf{c}) = e\sum_{i \in N} c_i$ and split equally amongst the players (if $e < 1$, $R(\mathbf{c})$ would be a 'public bad'). Write $\boldsymbol{\phi}$ for the payoff vector $\{\phi_1, \ldots, \phi_n\}$. Given others' contributions $\mathbf{c}_{-i}$, player i's contribution c_i results in the payoff

$$\phi_i = \frac{e}{n}\sum_{i \in N} c_i + (B - c_i).$$

2.3. Nash equilibria

If the rate of return is *low* ($e < n$), an individual contribution of zero ('free-riding') is the strictly dominant strategy for all players. Similarly, if the rate of return is *high* ($e > n$), an individual contribution of B ('fully contributing') is the strictly dominant strategy for all players. The respective Nash equilibria result in either nonprovision or full provision of the public good.

2.4. Repeated game

In each experimental session, the same population S (with $|S| = 12$ or 16) plays four 'phases' where each phase is a twenty-times repeated voluntary contribution game. In each period t within a given phase, players in S are randomly matched (as in Andreoni (1988)) in groups of four to play the voluntary contribution game. At the start of each period, each subject is given a new budget $B = 40$, of which he can invest any amount; however, subjects cannot invest money carried over from previous periods. Other than that, subjects are free to invest any number of coins between 0 and 40 each periods.

The rate of return is either *low* ($e = 1.6$) or *high* ($e = 6.4$) throughout a given phase. Write N_4^t for any of the four-player groups matched at time t. Note that, due to random group rematching, for every player i, the relevant group N_4^t (i.e. such that $i \in N_4^t$ in period t) typically is a different one in each period t. Given others' contributions $\mathbf{c}_{-i}^t$ for all t, each i receives a total of $\phi_i = \sum_{t=1,\ldots,20} \phi_i^t$. The number ϕ_i represents a monetary reward that is paid after the game.

2.5. Information treatments

Each experimental session (involving four phases) is divided into two 'stages': phases one and two of the session constitute stage one, phases three and four constitute stage two. Each phase is a twenty-period voluntary contribution game with either the low ($e = 1.6$) or the high ($e = 6.4$) rate of return, and in each stage both rates of return are played. The two stages differ with respect to the information revealed in that treatment.

Before the experiment begins, subjects are told that two separate experiments will be conducted, each stage consisting of two games. At no point before, during, or in between the two separate stages of the experiment are players allowed to communicate. Depending on which treatment is played, the following information is revealed at the start of each stage.[7]

All treatments. During each phase, the same game is repeated for 20 periods. Each player receives 40 coins each period, of which he can invest any amount. The amount not invested goes straight into his private account each period. After investments are made, each player earns a nonnegative return from his investment each period which, at the end of the experiment, he receives together with his uninvested money according to a known exchange rate into real money.
Black box. Subjects have no information about the structure of the game or about other players' actions and payoffs. Subjects play two voluntary contribution games (one with the high and one with the low rate of return). As play proceeds, subjects only know their own contributions and realized payoffs.
Standard. The rules of the game are revealed, including production of the public good, high or low rate of return, and how groups form each period. As the game proceeds, players receive a summary of the relevant contributions in their group at the end of each period.
Enhanced. In addition to the information available in the standard treatment, the payoffs of the other group members are explicitly calculated, and players receive a summary of their own and other players' payoffs in their group at the end of each period.

In each experimental session, every player plays two black box games (one with the high and one with the low rate of return), and either two standard or two enhanced information games (again one high and one low). Either the black box treatment occurs first, or a non-black box treatment occurs. Of the total of 18,880 observations, 9440 are black box, 4640 are standard, and 4800 are enhanced. The order, of high and low rates of return and of treatments, is different in each session. Sessions lasted between fifty and sixty minutes, and subjects earned between £6.20 and £15.50 (mean earnings were £12.40).

2.6. Black box details

Black box is our main treatment, and our main analysis is based on sessions when black box is played first. Recall that we require the recruiting system (ORSEE) to select only those who had not previously participated in public goods experiments. Thus, when black box is played first, subjects are not likely to have prior knowledge of the structure of the game.

> *We shall reserve the term 'black box' for the case when the 'no information' treatment is played first. Sessions when it is played after the standard or enhanced treatments will be called 'grey box'.*

In the 'black box' treatment, subjects were told that the black box would convert their input (in coins) to an output (in coins) via a mathematical function with a random component. The instructions (see Appendix B) are silent on the question of whether this function might depend on the actions of others. Subjects were not told that they would be engaged in a game, but the instructions do not rule out this possibility either. The asymmetric search behaviors that we are testing are a component of learning rules that have been proposed for highly uncertain environments of this type.

In the 'grey box' treatment, subjects are told that a new and separate experiment will be conducted and that all information except for their own payoffs will be withheld. Since play in these sessions was preceded by two voluntary contribution games where subjects received full information about the structure of the game, they might (or might not) think that the grey box treatment has a similar structure. However, they will still be unable to infer others' contributions and the underlying rates of return from the information they receive. We make this distinction between black box and grey box because our payoff-based learning model applies most evidently in the (pure) black box environment where more sophisticated learning models cannot apply due to the complete absence of information about the structure and about others' actions. Nevertheless, we shall find that many of the same learning behaviors are present in both treatments.

3. The learning model

In this section, we present and test our behavioral hypotheses based on the analysis of the pure black box data, that is, those sessions when black box was played before the standard or enhanced treatments. There were eight such sessions and 4960 observations.

[7] See Appendix B for instructions. Further details about the experimental instructions such as screenshots can be found in Burton-Chellew and West (2013).

3.1. Five behavioral hypotheses

Formally, we say that there is "inertia" in period t, that is, between periods t and $t-1$, if $c_i^t = c_i^{t-1}$. Regarding non-inertial adjustments in period t, we distinguish between "experimentation", present after previous payoff patterns were steady or increasing ($\phi_i^{t-1} \geq \phi_i^{t-2}$), and "exploration", which is triggered by payoff decreases ($\phi_i^{t-1} < \phi_i^{t-2}$).[8]

Consider the following example to illustrate our five learning hypotheses. Suppose that over three successive periods, $t-2$, $t-1$ and t, an agent contributes $c_i^{t-2} = 10$, $c_i^{t-1} = 10$ and $c_i^t = x$ with realized payoffs ϕ_i^{t-2}, ϕ_i^{t-1} and ϕ_i^t respectively.

(*i*) *Asymmetric inertia.* $x = 10$ is less likely if $\phi_i^{t-1} < \phi_i^{t-2}$ than if $\phi_i^{t-1} \geq \phi_i^{t-2}$.

(*ii*) *Asymmetric volatility.* The variance of x is higher if $\phi_i^{t-1} < \phi_i^{t-2}$ than if $\phi_i^{t-1} \geq \phi_i^{t-2}$.

(*iii*) *Asymmetric breadth.* The expected value of $|x-10|$ is higher if $\phi_i^{t-1} < \phi_i^{t-2}$ than if $\phi_i^{t-1} \geq \phi_i^{t-2}$.

(*iv*) *Reversion.* Suppose that $x \neq 10$ and that $c_i^{t+1} = y$ is chosen in the next period ($t+1$). In period $t+1$, $y = 10$ is more likely if $\phi_i^t < \phi_i^{t-1}$ than if $\phi_i^t \geq \phi_i^{t-1}$.

(*v*) *Directional bias.* Suppose that $x \neq 10$ and that $c_i^{t+1} = y$ is chosen in period $t+1$. If $x > 10$, then, in expectation, $y - x$ is larger when $\phi_i^t \geq \phi_i^{t-1}$ than when $\phi_i^t < \phi_i^{t-1}$. Similarly, if $x < 10$, then, in expectation, $y - x$ is smaller when $\phi_i^t \geq \phi_i^{t-1}$ than when $\phi_i^t < \phi_i^{t-1}$. In other words, the direction of successful adjustments is reinforced.

We find that all five features are confirmed at a statistically significant level (see Tables A1 and A2). The tests of significance were conducted as follows. To test the hypothesis of asymmetric inertia, we used an ordered probit regression of the inertia rate controlling for session, phase, group, period and individual fixed effects with individual-level clustering. Payoff decreases turn out to have a significant effect on the inertia rate: on average it reduces inertia (i.e. increases search probability) by 12% (± 2%): see Table A3.[9] There is significantly less inertia in the first phase than in the second phase, and period fixed effects are negative until period six, suggesting that inertia tends to increase over time. Importantly, asymmetric inertia is already present in the 'early' periods (before period 10) of the experiment, as well as in the 'later' periods (after period 10).

We also examine the situation in which a subject keeps the same action for two periods in a row and in the second period experiences a payoff shock due to a change of strategy by someone else (whom he knows nothing about due to the black box environment). Again, we find an asymmetric effect on search behavior in the next (third) period: on average a negative payoff shock increases the subsequent probability of search by 13% (± 4%) relative to a positive payoff shock. The baseline inertia rate in this setting however is roughly 50% higher (± 3%) than when a subject did not keep the same action for the two prior periods.

These findings are broadly consistent with Nevo and Erev (2012)'s 'surprise triggers change' hypothesis when payoff shocks are negative, but they are inconclusive when payoff shocks are positive. In the Nevo and Erev (2012) experiment, subjects either get very low payoffs (0, or −1) or very high payoffs (10 or −10), whereas in our setting, subjects experience a whole range of payoffs and there is no clear demarcation between 'ordinary' and 'surprise' payoffs. It is conceivable that if a subject happens to experience a series of small payoffs and then suddenly gets a very large increase, inertia might go down but our dataset does not contain the necessary number of cases to be able to test this possibility.

For asymmetric volatility, we apply Levene's test, a nonparametric test for the equality of variances in different samples. The null hypothesis of equal variances following steady/increasing payoffs versus decreasing payoffs, which would imply no asymmetric volatility, is rejected with 99% confidence. The same is true for early and late periods of the game. This holds for both high and low rates of return and for different orders of the games (see Table A4).

We test asymmetric breadth by regressing the absolute values of the non-inertial adjustments conditional on steady/increasing payoffs versus payoff declines, controlling for session, phase, group, period and individual fixed effects with individual-level clustering. The average adjustment following steady/increasing payoffs is smaller than after payoff declines (99% confidence; see Table A5). The same is true for early and late periods. Moreover, breadth is significantly larger in the first phase than in the second. Other fixed effects are not significant.

We test for reversion much as we did for asymmetric inertia. Payoff decreases turn out to have significantly positive effects on the reversion rate (see Table A6). Again, the same is true for early and late periods, but the effect is smaller in the first phase than in the second phase, i.e. reversion increases over time. Other fixed effects are not significant.

Finally, we analyze the directional patterns of adjustments. Our hypothesis of directional bias states that (i) if an increase leads to steady/increasing payoffs, the player's next-period contribution will tend to be higher than if it leads to payoff declines; (ii) if a decrease leads to steady/increasing payoffs, the player's contribution next period will tend to be lower than if it leads to payoff declines. To test this hypothesis, we regress the difference between adjustments after steady/increasing payoffs and adjustments after payoff declines, controlling for the direction of the previous adjustment as well as for session, phase, group, period and individual fixed effects with individual-level clustering. The directional bias is confirmed, for both prior contribution increases and prior contributions decreases, at high levels of significance ($p < 0.01$). The details are given in Table A7.

[8] The inclusion of stable payoffs in "experimentation" (i.e. not in "exploration") is a key characteristic of trial-and-error learning (Young, 2009).

[9] The range corresponds to the 95% confidence interval.

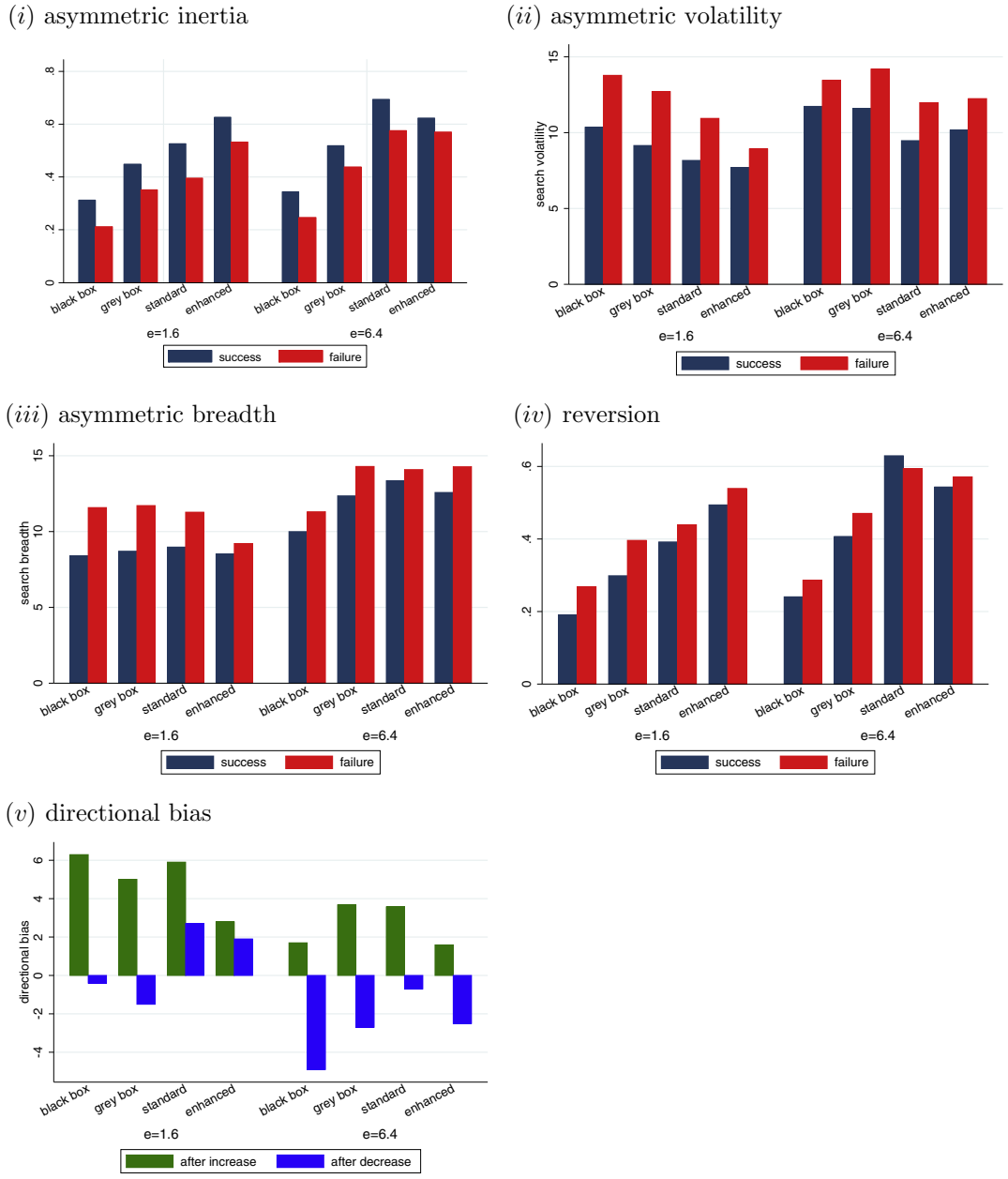


Fig. 1. All treatments, both rates of return. The bar charts summarize the five learning model components in all treatments for both rates of return separately. For (ii), (iv) and (v), the *y*-axes are units of contributions; for (i) and (iii), the *y*-axes are probabilities. We denote the cases of steady or increasing payoffs by 'success', and decreasing payoffs by 'failure'.

3.2. Non-black box learning

In this section, we assess our black box findings in light of the data from grey box and from the other two treatments, standard and enhanced. In particular, we investigate whether our learning model describes only black box behavior or whether its components also persist in the other settings, that is, where players gain experience and/or have explicit information about the structure of the game and others' actions (and payoffs).

Fig. 2 illustrates play in the different treatments, averaged over individuals and sessions. Different information clearly makes a difference in the level of contributions and/or convergence rates under the high rate of return (see also Burton-Chellew and West, 2013). However, the basic features of our learning model persist in all sessions and treatments, as is summarized in Fig. 1.

A noteworthy feature is that convergence to equilibrium is both more pronounced and more similar across treatments when the rate of return is *low* ($e = 1.6$) than when it is *high* ($e = 6.4$). Furthermore, contributions in the low-rate-of-return

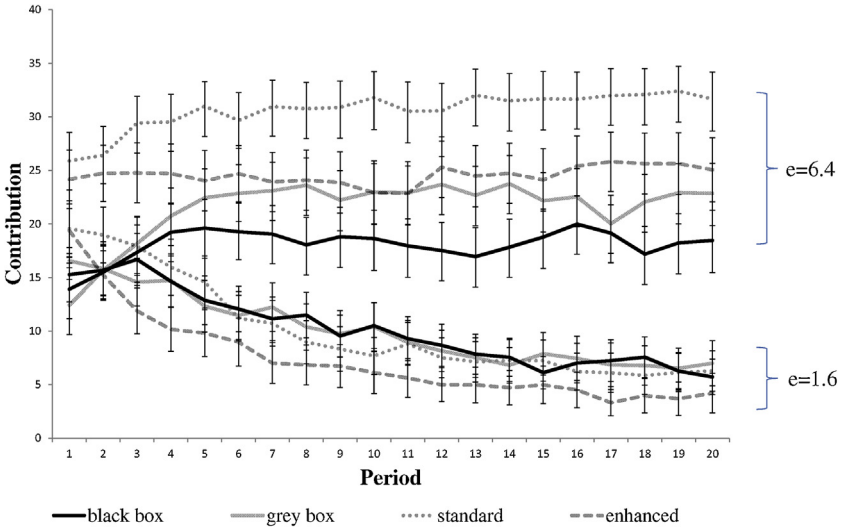


Fig. 2. Mean play all treatments (black box, grey box, standard, enhanced).

case are higher in the standard treatment than in the enhanced treatment. Under the high rate of return, contributions tend to polarize over time: some subjects appear to learn the dominant strategy (full contribution) whereas others do not. This contrasts with the learning dynamics under the low rate of return, where contributions gradually shifted from near-random contributions in early period to low-or-zero contributions in the later period.[10] Hence, providing more information can actually lead to decreasing play of the dominant strategy, which supports the hypothesis that our five behavioral features still have bite with more information. In other words, some subjects continue to act as if they were inside a black box even though they are not. The slow-down of convergence in the high-rate-of-return case, especially under black box, could be the result of the 'payoff variability effect' (Myers and Sadler, 1960; Busemeyer and Townsend, 1993; Erev and Barron, 2005; Erev and Haruvy, 2013), according to which increased payoff variability may inhibit learning (see Appendix A, Fig. 4). In the standard and enhanced conditions, it could moreover be the case that subjects did not understand (fully) that they had a dominant strategy, or that relative payoff concerns interfered with equilibrium play. We leave further investigations of this feature to future research.

3.2.1. *Black box versus grey box*

Next, we shall investigate 'grey box' play. Recall that grey box means that black box is played after either enhanced or standard. Even though subjects are explicitly told that a separate experiment is started after the first stage of the experiment, players may or may not make inferences about the game structure. We shall investigate the consequences of this effect in comparison with (pure) black box behavior. Our analysis reveals that, although there are differences in average levels of contribution in the two cases (Fig. 2), all of the features of our learning model are robust for both rates of return (see also Fig. 1). In particular, the differences in inertia rates conditional on steady-or-increasing payoff versus decreasing payoff are similarly robust, however, the level of inertia in absolute terms is higher in grey box than in black box (Test 1, Table A8). The levels of asymmetric volatility and asymmetric breadth are not significantly different in the two cases (Test 1, Tables A9 and A10). The differences in reversion are robust, but higher in grey box than in black box (Test 1, Table A11). Directional bias is unchanged in terms of size and significance (Test 1, Table A12). We conclude that grey box differs from black box with regard to the absolute levels of inertia and adjustment, but not qualitatively with respect to any of the five learning model components.

3.2.2. *Black box versus standard and enhanced*

Finally, we consider whether the standard and enhanced treatments lead to different conclusions. The situation is summarized in Fig. 1. Panel (i) in Fig. 1 illustrates that inertia rates decrease after payoff decreases in all treatments. This steady-or-increasing payoffs versus decreasing payoffs differential in inertia rates is statistically significant in all four treatments, and there is higher absolute inertia in non-black box treatment data (see Test 2, Table A8). Asymmetric volatility, though smaller in absolute size in the standard treatment and even smaller in the enhanced treatment, is higher after payoff decreases than after steady or increasing payoffs in all treatments (see panel (ii) in Fig. 1). Moreover, this difference is significant at the 1% level (see Test 2, Table A9). Asymmetric breadth, too, exhibits an analogous differential as in black box (see panel (iii) in Fig. 1). It continues to be statistically significant at the 1% level (see Test 2, Table A10). Panel (iv) in Fig. 1 illustrates that reversion rates increase after payoff decreases in all treatments except for under the standard treatment

[10] See Appendix A, Fig. 3 for initial versus final contributions in the different treatments.

when $e = 6.4$, but significant overall (see Test 2, Table A11). The difference is statistically significant in all four treatments, and there are higher absolute reversion rates in non-black box treatment data (see Test 2, Table A11). Finally, panel (v) in Fig. 1 illustrates the presence of directional biases across all four treatments. We qualitatively confirm the directional bias from the black box data when analyzing the non-black box data at levels that are statistically significant (see Test 2, Table A12).

4. Conclusion

Much of the prior empirical work on learning in games has focussed on situations where players have a substantial amount of information about the structure of the game and they can observe the behavior of others as the game proceeds. In this paper, by contrast, we have examined situations in which players have *no* information about the strategic environment. This takes us to the other end of the information spectrum. Players in such environments must feel their way toward equilibrium based solely on the pattern of their own realized payoffs.

To highlight the potential differences between adaptive learning and best-reply dynamics, we implemented a black box environment in a voluntary contribution game and compared the resulting behavior to play under intermediate and rich information treatments. We identified five key features of the learning dynamics: asymmetric inertia, asymmetric volatility, asymmetric breadth, reversion, and directional bias. Although these components have precursors in both psychology and biology, they have not been given the precise formulation that we propose here. It turns out all five features are validated at high levels of statistical significance in the black box treatment. Moreover, they are present even when players gain more experience and/or have more information about the game. Whether this remains true for other classes of games is an open question for future research.

Acknowledgements

We thank Margherita Comola, Johannes Abeler, Guillaume Hollard, Richard Bradley, Fabrice Etilé, Philippe Jehiel, Thierry Verdier, Vince Crawford, Colin Camerer, Muriel Niederle, Edoardo Gallo, Amnon Rapoport, Guillaume Fréchette, Ozan Aksoy, Margaret Meyer, Roberto Serrano, Arno Riedl, Martin Strobel, Martin Cripps, Madis Ollikainen, Stefano Duca, Aidas Masiliunas, anonymous referees, a helpful editor, and participants of the Economics and Psychology seminar at MSE Paris, the LSE Choice Group, the 24th Game Theory Workshop at Stony Brook, the CESS seminar at Nuffield College, the 9th Tinbergen Institute Conference and the Economics seminar at the University of Maastricht for helpful discussions, suggestions and comments.

Appendix A. Regression outputs

In the following, a single star indicates $p < 0.05$, double-stars indicate $p < 0.01$.

Table A1
Summary statistics for asymmetric inertia, asymmetric volatility, asymmetric breadth, and reversion (black box).

(a) Learning component	(b)	(c) Payoff decreases	(d) Steady/increasing payoffs	(e) Difference (c) – (d)
(i) *asymmetric inertia*				
Probability of zero adjustments	Relative frequency	0.23	0.33	0.10**
	# observations	520	722	
(ii) *Asymmetric breadth*				
Absolute value of non-zero adjustments	Mean	11.4	9.2	2.2**
	# observations	1746	1476	
(iii) *Asymmetric volatility*				
Standard deviation of adjustment	Standard deviation	15.6	13.6	2.0**
	# observations	1746	1476	
(iv) *Reversion*				
Probability of return after search	Relative frequency	0.21	0.08	0.13**
	# observations	160	57	

Table A2
Summary statistics for (v) directional bias (black box).

(a) Mean non-zero adjustment	(b) Steady/increasing payoffs	(c) Payoff decreases	(d) Difference (b) – (c)
After *increase*	−1.6	−6.5	4.9**
After *decrease*	2.5	5.5	−3.0**

Table A3
Asymmetric inertia (black box). We use black box data from phases 1 and 2. We perform an ordered probit regression of absolute adjustments controlling for session, phase, group, period and individual fixed effects with individual-level clustering. The pseudo-R^2 indicates the level of improvement over the intercept model (not interpretable as percentage of variance explained).

	Coefficient (test statistic)
1 if "payoff decreases"	−0.52 (−10.39)**
1 if "period" < 10	0.22 (1.37)
1 if "phase" = 1	−1.21 (−50.38)**
Periods	Not listed
Individual fixed effects	Not listed
Group fixed effects	Not listed
Session fixed effects	Not listed
Cut	−0.05 (s.e. 0.13)
Observations	4464
Adjusted pseudo-R^2	0.40

Table A4
Asymmetric volatility (black box). We use black box data from phases 1 and 2. We perform Levene's robust variance test, W being the Levene's test statistic.

(a) Test 1: Asymmetric volatility Null rejected ($W = 39.9$, $p < 0.01$)	(b) Steady/increasing payoffs	(c) Payoff decreases	(d) Total
Standard deviation	13.6	15.6	14.8
Mean	+1.2	−1.5	−0.3
Frequency	1476	1746	3222

Table A5
Asymmetric breadth (black box). We use black box data from phases 1 and 2. We perform an OLS regression of absolute adjustments controlling for session, phase, group, period and individual fixed effects with individual-level clustering.

	Coefficient (test statistic)
1 if "payoff decreases"	2.28 (6.89)**
1 if "period" < 10	1.28 (1.25)
1 if "phase" = 1	2.75 (2.40)*
Periods	Not significant
Individual fixed effects	Not listed
Group fixed effects	Not listed
Session fixed effects	Not listed
Constant	4.49 (3.07)**
Observations	3222
Adjusted R^2	0.30

Table A6
Reversion (black box). We use black box data from phases 1 and 2. We perform an ordered probit regression of reversion rates controlling for session, phase, group, period and individual fixed effects with individual-level clustering. The pseudo-R^2 indicates the level of improvement over the intercept model (not interpretable as percentage of variance explained).

	Coefficient (test statistic)
1 if "payoff decreases"	0.72 (8.39)**
1 if "period" < 10	−0.42 (−1.73)*
1 if "phase" = 1	0.08 (1.37)
Periods	Not significant
Individual fixed effects	Not listed
Group fixed effects	Not listed
Session fixed effects	Not listed
Cut	1.29 (s.e. 0.19)
Observations	3252
Adjusted pseudo-R^2	0.26

Table A7
Directional bias (black box). We use black box data from phases 1 and 2. We perform OLS regressions (without constant) to test for the directional bias of adjustments for each directional impulse controlling for phase, group, period and individual fixed effects with individual-level clustering.

	Coefficient (test statistic)
1 if "up"	−3.48 (1.14)
1 if "up" and "steady or increasing payoff"	4.53 (5.35)**
1 if "down"	9.20 (2.83)**
1 if "down" and "steady or increasing payoff"	−3.07 (3.70)**
1 if "period"<10	−0.25 (−0.19)
Periods	Not significant
Individual fixed effects	Not listed
Group fixed effects	Not listed
Phase fixed effects	Not listed
Session fixed effects	Not listed
Observations	3264
Adjusted R^2	0.11

Table A8
Asymmetric inertia (non-black box). We use non-black box data; phases 3 and 4 for grey box (*Test 1*), and phases 1–4 for standard and enhanced (*Test 2*). We perform an ordered probit regression of absolute adjustments controlling for treatment effects as well as session, phase, group, period and individual fixed effects with individual-level clustering. The pseudo-R^2 indicates the level of improvement over the intercept model (not interpretable as percentage of variance explained).

	Test 1: Grey box coefficient (test statistic)	**Test 2: Standard &enhanced** coefficient (test statistic)
1 if "payoff decreases"	−0.45 (−9.14)**	−0.34 (9.98)**
1 if "period"<10	−0.49 (−3.17)**	−0.35 (3.15)**
1 if "enhanced"	n/a	0.96 (44.85)**
Periods	Not significant	Not significant
Other fixed effects	Not listed	Not listed
Cut	5.05 (s.e. 0.23)	0.52 (s.e. 0.09)
Observations	4032	8464
Adjusted pseudo-R^2	0.39	0.44

Table A9
Asymmetric volatility (non-black box). We use non-black box data; phases 3 and 4 for grey box, phases 1–4 for standard and enhanced. We perform Levene's robust variance tests for asymmetric volatility in grey box (Test 1), and standard and enhanced (Test 2). Recall W is the Levene's test statistic.

	Steady/increasing payoffs	Decreasing payoffs	Total
Test 1: *Grey box*			
Null rejected ($W = 23.3$, $p < 0.01$)			
Standard deviation	14.6	17.3	16.2
Mean	+2.0	−1.9	−0.1
Frequency	1053	1206	2259
Test 2: *Standard and enhanced*			
Null rejected ($W = 9.3$, $p < 0.01$)			
Standard deviation	14.4	15.8	15.5
Mean	+2.8	−3.6	−0.6
Frequency	1707	1931	3638

Table A10
Asymmetric breadth (non-black box). We use non-black box data; phases 3 and 4 for grey box (*Test 1*), phases 1–4 for standard and enhanced (*Test 2*). We perform OLS regressions of absolute adjustments controlling for session, phase, group, period and individual fixed effects with individual-level clustering.

	Test 1: Grey box coefficient (test statistic)	Test 2: Standard and enhanced coefficient (test statistic)
1 if "payoff decreases"	2.09 (5.10)**	1.36 (4.93)**
1 if "period"<10	Not significant	Not significant
Periods	Not significant	Not significant
Other fixed effects	Not listed	Not listed
Constant	19.91 (16.71)**	18.85 (19.68)**
Observations	2259	3638
Adjusted R^2	0.42	0.47

Table A11
Reversion (non-black box). We use non-black box data; phases 3 and 4 for grey box (*Test 1*), and phases 1–4 for standard and enhanced (*Test 2*). We perform an ordered probit regression of absolute adjustments controlling for treatment effects as well as session, phase, group, period and individual fixed effects with individual-level clustering. The pseudo-R^2 indicates the level of improvement over the intercept model (not interpretable as percentage of variance explained).

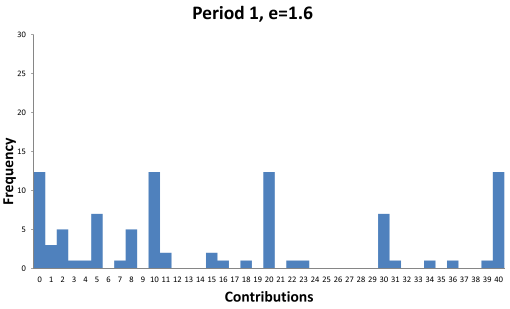
	Test 1: Grey box coefficient (test statistic)	Test 2: Standard &Enhanced coefficient (test statistic)
1 if "payoff decreases"	0.82 (8.17)**	0.60 (8.70)**
1 if "enhanced"	n/a	−0.58 (−11.08)**
1 if "period"<10	−0.04 (0.15)	0.00 (−0.01)
Periods	Not significant	Not significant
Other fixed effects	Not listed	not listed
Cut	2.19 (s.e. 0.24)	1.01 (s.e. 0.17)
Observations	2304	3725
Adjusted pseudo-R^2	0.26	0.22

Table A12
Directional bias (non-black box). We use all the data. We perform OLS regressions (without constant) to test for the directional bias of adjustments for each directional impulse in grey box (Test 1) and standard and enhanced (Test 2) controlling for session, phase, group, period and individual fixed effects with individual-level clustering.

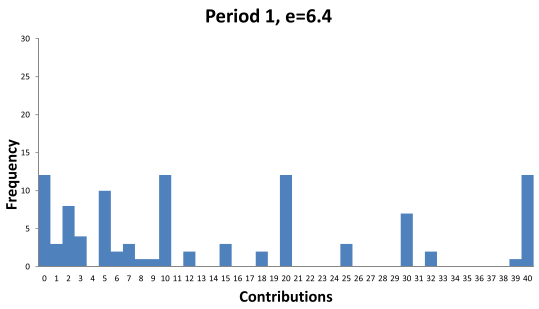
Test 1: Grey box	Adjustment: coefficient (test statistic)
1 if "up"	−3.86 (−1.88)*
1 if "up" and "steady or increasing payoffs"	5.24 (5.58)**
1 if "down"	8.91 (4.50)**
1 if "down" and "steady or increasing payoffs"	−2.41 (2.98)**
1 if "period"< 10	0.48 (0.31)
Period fixed effects	Not significant
Group dummies	Not significant
Other fixed effects	Not listed
Observations	2381
Adjusted R^2	0.18

Test 2: Standard and enhanced	Adjustment: coefficient (test statistic)
1 if "up"	−3.04 (−1.84)*
1 if "up" and "steady or increasing payoffs"	3.98 (5.53)**
1 if "down"	8.16 (4.76)**
1 if "down" and "steady or increasing payoffs"	−0.50 (−0.68)
1 if "enhanced"	−1.30 (−0.88)
1 if "period"<10	0.32 (0.27)
Period fixed effects	Not significant
Group dummies	Not significant
Other fixed effects	Not listed
Observations	3938
Adjusted R^2	0.22

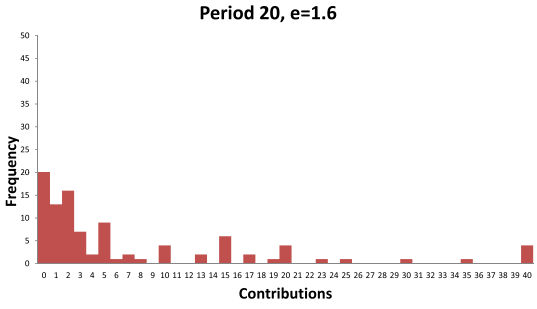
(*i*) period 1, low rate



(*ii*) period 1, high rate



(*iii*) period 20, low rate



(*iv*) period 20, high rate

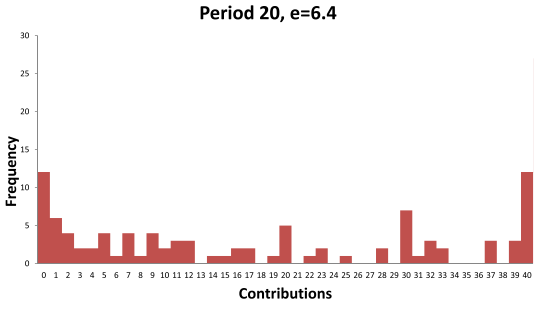


Fig. 3. Initial and final contributions for both rates of return under black box.

Appendix B. Instructions of the lab experiment

Participants received the following on-screen instructions (in z-Tree) at the start of the game. The set of instructions in standard and enhanced were the same, different instructions were given in black box. In black box, participants had to click an on-screen button saying, "I confirm I understand the instructions" before the game would begin. The same black box instructions were used for both rates of return. The standard/enhanced instructions differ with respect to the relevant numbers for the two rates of return, and the example is adequately modified.

B.1. Black box

The following instructions were used in black box.

Beginning of instruction.

Instructions

Welcome to the experiment. You have been given 40 virtual coins. Each 'coin' is worth real money. You are going to make a decision regarding the investment of these 'coins'. This decision may increase or decrease the number of 'coins' you have. The more 'coins' you have at the end of the experiment, the more money you will receive at the end.

During the experiment we shall not speak of £ Pounds or Pence but rather of "Coins". During the experiment your entire earnings will be calculated in Coins. At the end of the experiment the total amount of Coins you have earned will be converted to Pence at the following rate: 100 Coins = 15 Pence. In total, each person today will be given 3200 coins (£4.80) with which to make decisions over 2 economic experiments and their final totals, which may go up or down, will depend on these decisions.

The Decision

You can choose to keep your coins (in which case they will be 'banked' into your private account, which you will receive at the end of the experiment), or you can choose to put some or all of them into a '**black box**'.

This '**black box**' performs a mathematical function that converts the number of coins inputted into a number of coins to be outputted. The function contains a random component, so if two people were to put the same amount of coins into the '**black box**', they would not necessarily get the same output. The number outputted may be more or less than the number you put in, but it will never be a negative number, so the lowest outcome possible is to get 0 (zero) back. If you chose to input 0 (zero) coins, you may still get some back from the box.

Any coins outputted will also be 'banked' and go into your private account. So, your final income will be the initial 40 coins, minus any you put into the '**black box**', plus all the coins you get back from the '**black box**'.

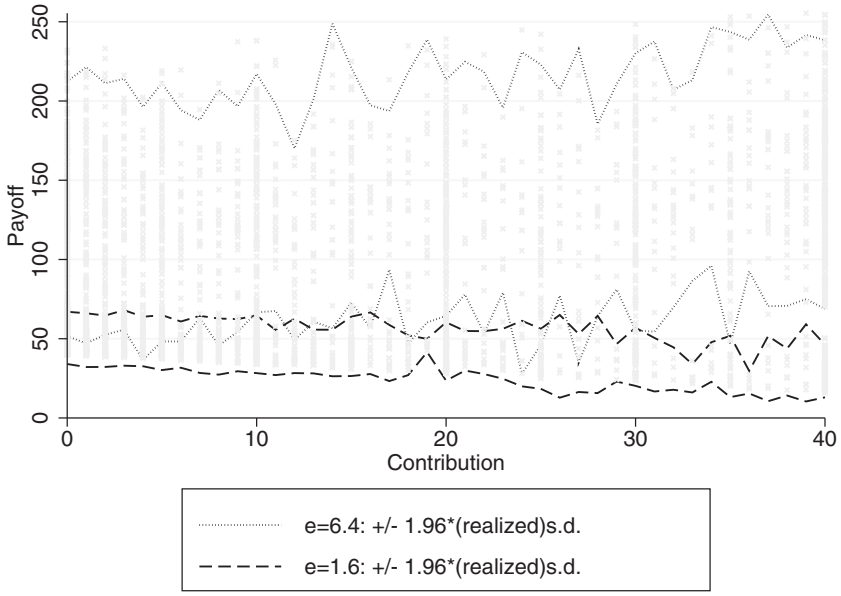


Fig. 4. Payoff variability (black box).

You will play this game 20 times. Each time you will be given a new set of 40 coins to use. Each game is separate but the '**black box**' remains the same. This means you cannot play with money gained from previous turns, and the maximum you can ever put into the '**black box**' will be 40 coins. And you will never run out of money to play with as you get a new set of coins for each go. The mathematical function will not change over time, so it is the same for all 20 turns. However as the function contains a random component, the output is not guaranteed to stay the same if you put the same amount in each time.

After you have finished your 20 turns, you will play one further series of 20 turns but with a new, and potentially different '**black box**'. The two boxes may or may not have the same mathematical function as each other, but the functions will always contain a random component, and the functions will always remain the same for the 20 turns. You will be told when the 20 turns are finished and it is time to play with a new black box.

If you are unsure of the rules please hold up your hand and a demonstrator will help you.

I confirm I understand the instructions (click to confirm)

End of instructions.

B.2. Standard and enhanced

Here, we present the instructions for the rate of return $e = 1.6$. The same instructions apply to standard and enhanced treatments. Equivalent instructions apply for the rate of return $e = 6.4$.

Beginning of instructions.

Instructions

Welcome! You are about to participate in an experimental study of human decision making. Thank you for your participation in our study. Please pay careful attention to the instructions on the following screens. If you wish to return to a previous screen, press the left arrow key. You are now taking part in an economic experiment. If you read the following instructions carefully, you can, depending on your decisions, earn a considerable amount of money. It is therefore very important that you read these instructions with care.

Everybody has received the same instructions. It is prohibited to communicate with the other participants during the experiment. Should you have any questions please ask us by raising your hand. If you violate this rule, we shall have to exclude you from the experiment and from all payments.

During the experiment we shall not speak of £ Pounds or Pence but rather of "Credits". During the experiment your entire earnings will be calculated in Credits. At the end of the experiment the total amount of Credits you have earned will be converted to Pence at the following rate: 100 Credits = 15 Pence. In total, each person today will be given 3200 credits (£4.80) with which to make decisions over two economic experiments and their final totals, which may go up or down, will depend on these decisions.

We are researching the decisions people make.

This part of the experiment is divided into separate rounds. In all, this part of the experiment consists of 20 repeated rounds. In each round the participants are assembled into groups of four. You will therefore be in a group with 3 other

participants. The composition of the groups will change at random after each round. **In each round your group will therefore probably consist of different participants.**

In each round the experiment consists of two stages. At the first stage everyone has to individually decide how many credits they would like to contribute to a group project. These decisions have consequences for people's earnings. At the second stage you are informed of the contributions of the three other group members to the project and how many credits you have received from the group project. New groups are then randomly formed and the process repeats itself, with everyone again deciding how many credits they would like to contribute. This process will repeat 20 times.

At the beginning of each round, each participant receives 40 credits. In the following we call this your endowment. Your task is to decide how to use your endowment. You have to decide how many of the 40 credits you want to contribute to a group project and how many of them to keep for yourself. The consequences of your decision are explained in detail in the following slides.

Please note:

- The set-up is anonymous. You will not know with whom you are interacting.
- You will interact with a random set of 3 players in each round.
- Your decisions, and your earnings will remain anonymous to other players, even after the session has ended.

We now provide an animated illustration of a hypothetical scenario to demonstrate how the experiment works. In the following demonstration, we will use these green 'disks' to represent 'credits'. 1 disk equals 1 credit.

Example

Remember! Credits = real money, and the more credits a player has at the end of the experiment, the more money that player will receive.

There are 4 players in each group, For the sake of convenience, we refer to these players as Player A, Player B, Player C, and Player D.

These 4 players then each receive an endowment of 40 credits.

Each of the players can then choose to make a contribution to the group project. They can contribute anything from zero to 40 credits. Non contributed credits are kept in the player's private account. They do this at the same time and anonymously.

Each player is then informed of the decisions of all their group members, although no one will know who the players are and they will randomly change in each round.

After all 4 players have made their decision to contribute or not, and by how much, the resulting total of contributed credits is automatically MULTIPLIED.

In your experiment, in every round, and in this example here, the total will be multiplied by 1.6. So for an imaginary total of 10 credits, this would result in 16 credits.

This new total is then always shared out equally between all 4 players. So, after the multiplication occurs, each player receives one quarter of the credits that are in the group project.

In this example, each player chooses to contribute 20 of their 40 credits. This means the group project has 80 credits; 20 + 20 + 20 + 20 = 80 credits. Which when multiplied by 1.6 results in 128 credits; 80 multiplied by 1.6 = 128 credits. These 128 credits are then shared out equally, giving 32 credits back to each player; 128 divided by 4 = 32 credits each. This gives each of the players a new total. In this case, they all have a new total of 52 credits.

They all started with +40; Contributed −20; and all got +32 in return, **giving them 52 in total.**

That is the end of the demonstration.

Remember, this was just one of many possible scenarios. In the rounds you will now play, all players are free to choose how much they wish to contribute to the pot.

End of instructions.

References

Andreoni, J., 1988. Why free ride? Strategies and learning in public goods experiments. J. Public Econ. 37, 291–304.
Bayer, R.-C., Renner, E., Sausgruber, R., 2013. Confusion and learning in the voluntary contributions game. Exp. Econ. 16, 478–496.
Ben Zion, U., Erev, I., Haruvy, E., Shavit, T., 2010. Adaptive behavior leads to under-diversification. J. Econ. Psychol. 31, 985–995.
Bowling, M., Veloso, M., 2002. Multiagent learning using a variable learning rate. Artif. Intell. 136 (2), 215–250.
Burton-Chellew, M.N., West, S.A., 2013. Pro-social preferences do not explain human cooperation in public-goods games. Proc. Natl. Acad. Sci. 110 (31), 216–221.
Burton-Chellew, M.N., Nax, H.H., West, S.A., 2015. Payoff-based learning explains the decline in cooperation in public goods games. Proc. R. Soc. Lond. B: Biol. Sci. 1801 (282).
Busemeyer, J.R., Townsend, J.T., 1993. Decision field theory: a dynamic-cognitive approach to decision making in an uncertain environment. Psychol. Rev. 100, 432–459.
Bush, R.R., Mosteller, F., 1953. A stochastic model with applications to learning. Ann. Math. Statist. 24 (4), 559–585.
Chaudhuri, A., 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. Exp. Econ. 14, 47–83.
Colman, A.M., Pulford, B.D., Omtzigt, D., al Nowaihi, A., 2010. Learning to cooperate without awareness in multiplayer minimal social situations. Cognit. Psychol. 61 (3), 201–227.
Cross, J.G., 1983. A Theory of Adaptive Economic Behavior. Cambridge University Press.
Eiben, A.E., Schippers, C.A., 1998. On evolutionary exploration and exploitation. Fundam. Inf. 35 (1), 35–50.
Erev, I., Barron, G., 2005. On adaptation, maximization, and reinforcement learning among cognitive strategies. Psychol. Rev. 112, 912–931.
Erev, I., Haruvy, E., 2013. Learning and the economics of small decisions. Handb. Exp. Econ. 47, 501–512.

Erev, I., Rapoport, A., 1998. Coordination, 'magic', and reinforcement learning in a market entry game. Games Econ. Behav. 23 (2), 146–175.
Foster, D.P., Young, H.P., 2006. Regret testing: learning to play Nash equilibrium without knowing you have an opponent. Theor. Econ. 1, 341–367.
Friedman, D., Huck, S., Oprea, R., Weidenholzer, S., 2015. From imitation to collusion: long-run learning in a low-information environment. J. Econ. Theory 155 (C), 185–205.
Germano, F., Lugosi, G., 2007. Global Nash convergence of Foster and Young's regret testing. Games Econ. Behav. 60, 135–154.
Harley, C.B., 1981. Learning the evolutionarily stable strategy. J. Theor. Biol. 89 (4), 611–633.
Harstad, R.M., Selten, R., 2013 June. Bounded-rationality models: tasks to become intellectually competitive. J. Econ. Lit. 51 (2), 496–511.
Hart, S., Mas-Colell, A., 2003. Uncoupled dynamics do not lead to Nash equilibrium. Am. Econ. Rev. 93, 1830–1836.
Hart, S., Mas-Colell, A., 2006. Stochastic uncoupled dynamics and Nash equilibrium. Games Econ. Behav. 57, 286–303.
Herrnstein, R.J., 1970. On the law of effect. J. Exp. Anal. Behav. 13, 243–266.
Huttegger, S.M., Skyrms, B., 2012. Emergence of a signaling network with "probe and adjust. In: Signaling, Commitment, and Emotion. MIT Press, Cambridge, MA.
Huttegger, S.M., Skyrms, B., Zollman, K.J.S., 2014. Probe and adjust in information transfer games. Erkenntnis 79 (4), 835–853.
Isaac, M.R., McCue, K.F., Plott, C.R., 1985. Public goods provision in an experimental environment. J. Public Econ. 26, 51–74.
Isaac, R.M., Walker, J.M., 1988. Group size effects in public goods provision: the voluntary contributions mechanism. Q. J. Econ. 103 (1), 179–199.
Ledyard, J.O., 1995. Public goods: a survey of experimental research. In: Kagel, J.H., Roth, A.E. (Eds.), In: Handbook of Experimental Economics, vol. 37, pp. 111–194.
Marden, J.R., Young, H.P., Arslan, G., Shamma, J.S., 2009. Payoff-based dynamics for multiplayer weakly acyclic games. SIAM J. Control Optim. 48 (1), 373–396.
Marden, J.R., Young, H.P., Pao, L.Y., 2014. Achieving Pareto optimality through distributed learning. SIAM J. Control Optim. 52 (5), 2753–2770.
Motro, U., Shmida, A., 1995. Near-far search: an evolutionary stable foraging strategy. J. Theor. Biol. 173, 15–22.
Myers, J.L., Sadler, E., 1960. Effects of range of payoffs as a variable in risk taking. J. Exp. Psychol. 60, 306–309.
Nax, H.H., Perc, M., 2015. Directional learning and the provisioning of public goods. Sci. Rep. 5, 8010.
Nevo, I., Erev, I., 2012. On surprise, change, and the effect of recent outcomes. Front. Cognit. Sci. 47, 501–512.
Nowak, M., Sigmund, K., et al., 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. Nature 364 (6432), 56–58.
Pradelski, B.S.R., Young, H.P., 2012. Learning efficient Nash equilibria in distributed systems. Games Econ. Behav. 75, 882–897.
Rapoport, A., Seale, D.A., Parco, J.E., 2002. Coordination in the aggregate without common knowledge or outcome information. In: Experimental Business Research. Springer, USA, pp. 69–99.
Robbins, H., 1952. Some aspects of the sequential design of experiments. Bull. Am. Math. Soc. 58 (5), 527–535.
Roth, A.E., Erev, I., 1995. Learning in extensive-form games – experimental data and simple dynamic models in the intermediate term. Games Econ. Behav. 8, 164–212.
Selten, R., 1998. Aspiration adaptation theory. J. Math. Psychol. 42, 191–214.
Selten, R., Stoecker, R., 1986. End behavior in sequences of finite prisoner's dilemma supergames: a learning theory approach. J. Econ. Behav. Organ. 7, 47–70.
Skyrms, B., 2010. Signals: Evolution, Learning, and Information. Oxford University Press.
Skyrms, B., 2012. Learning to signal with probe and adjust. Episteme 9 (02), 139–150.
Suppes, P., Atkinson, A.R., 1959. Markov Learning Models for Multiperson Situations. Stanford University Press.
Thorndike, E.L., 1898. Animal Intelligence: An Experimental Study of the Associative Processes in Animals. Macmillan, New York.
Thuijsman, F., Peleg, B., Amitai, M., Shmida, A., 1995. Automata, matching and foraging behavior of bees. J. Theor. Biol. 175 (3), 305–316.
Weber, R.A., 2003. 'Learning' with no feedback in a competitive guessing game. Games Econ. Behav. 44, 134–144.
Young, H.P., 2009. Learning by trial and error. Games Econ. Behav. 65, 626–643.