Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2013/2

On the Impulse in Impulse Learning

Jieyao Ding
Andreas Nicklisch

Preprints of the
Max Planck Institute
for Research on Collective Goods Bonn 2013/2

On the Impulse in Impulse Learning

Jieyao Ding and Andreas Nicklisch

January 2013

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
http://www.coll.mpg.de

On the Impulse in Impulse Learning*

Jieyao Ding† and Andreas Nicklisch‡

January 29, 2013

Abstract

This paper experimentally investigates the nature of impulses in impulse learning. Particularly, we analyze whether positive feedback (i.e., yielding a superior payoff in a game) or negative feedback (i.e., yielding an inferior payoff in a game) leads to a systematic change in the individual choices. The results reveal that subjects predominantly learn from negative feedback.

Keywords: *Aspiration level, Impulse, Learning, Reinforcement, Stimulus*
JEL-Classification: C91 (Laboratory, Individual Behavior); D03 (Behavioral Economics); D83 (Learning).

*We gratefully acknowledge the many useful comments provided by Christoph Engel, Sven Fischer, Sebastian Goerg, Werner Güth, and Oliver Kirchkamp. We are indebted to the Max Planck Society for financial support.

†Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str.10, 53113 Bonn, Germany, Email: ding@coll.mpg.de

‡Corresponding author; University of Hamburg, School of Business, Economics and Social Science, von-Melle-Park 5, 20146 Hamburg, Germany, & Max Planck Institute for Research on Collective Goods, Email: andreas.nicklisch@wiso.uni-hamburg.de

1 Introduction

One of the big issues in behavioral economics is learning. The question of whether and (if so) how people's decisions converge towards an equilibrium by repeatedly playing is subject of a large number of studies (cf. the recent survey by Erev & Haruvy, 2012). One central approach is impulse learning. Although impulse learning models follow quite different ideas on how to model the attraction of a pure strategy in period t based on the experience of $t-1$ past periods, they share an important feature: receiving feedback about the (potentially) earned payoff by choosing strategy k in the period $t-1$ (the impulse) influences the confidence with which k is chosen in the current period. Prominent examples for impulse learning models are reinforcement learning (e.g., Roth & Erev, 1995, Erev & Roth, 1998), experience weighted attraction learning (e.g., Camerer & Ho, 1999, Ho et al., 2008), regret-based learning (e.g., Marchiori & Warglien, 2008), and learning (e.g., Chmura et al., 2012).[1]

In this study, we want to analyze the nature of the impulse in greater detail. For this purpose, we run a series of laboratory experiments on a simple two-player game, which allows us to analyze whether the impulse results from negative feedback or from positive feedback. That is, in the first case subjects increase the attraction of those actions whose choices yielded superior outcomes previously (reinforcement learning and in some sense experience attraction learning follow this idea), while in the second case subjects decrease the attraction of those actions whose choices yielded inferior outcomes previously (impulse matching learning and regret based learning propagate this view).[2] Both cases coincide for games with two pure actions for each player: the negative stimulus for one pure action equals the equivalent positive stimulus for the alternative action. However, considering a game with three pure actions allows us to disentangle the two potential sources for the impulse.

In contrast to previous studies (e.g., Grosskopf, 2003, and Chmura et al., 2012), we try to measure single impulses for feedback as narrowly as possible. For this purpose, we will eliminate any effect of aggregation over several periods of play, but observe the impulse in a game that is played twice only. As a consequence, our results are neither based on mean impulses resulting from a number of t interactions nor based on a single impulse, which, however, reflects the fact that a number of $t-2$ interactions are still to come.

[1]Complementary equilibrium concepts are provided by Beggs (2005) for reinforcement learning, and Selten & Chmura (2008) for impulse-matching learning (Ockenfels & Selten, 2005, and Selten et al., 2005, apply the same concept to repeated auctions).

[2]We will clarify the terms superior and inferior payoffs below.

2 The Game

We consider the following simultaneous-move game for two players. One of them, A, has three alternative actions {Up, Left, Right}; the other, B, is endowed with 100 points. B has to allocate the points on three options, $\{x, y, 100 - x - y\}$. Denoting A's choice as δ_A, payoff functions π_A and π_B for A and B, respectively, are

$$\pi_A = \begin{cases} 0.5x + 2y + (100 - x - y) & \text{if } \delta_A = \text{Up}, \\ 0.5x + 0.5y + (100 - x - y) & \text{if } \delta_A = \text{Left}, \\ 2x + y + 1.5(100 - x - y) & \text{if } \delta_A = \text{Right}, \end{cases} \tag{1}$$

and

$$\pi_B = \begin{cases} x + 1.5y + 0.5(100 - x - y) & \text{if } \delta_A = \text{Up}, \\ 2x + y + 0.5(100 - x - y) & \text{if } \delta_A = \text{Left}, \\ x + 0.5y + 2(100 - x - y) & \text{if } \delta_A = \text{Right}. \end{cases} \tag{2}$$

Another interpretation of the game is the following: both players face 100 identical games. A has to decide uniformly for all 100 games, while B can choose independently for each of the 100 games where to allocate one point. Obviously, Left is dominated for A, whereas $x = 100$ maximizes the minimum payoff for B. The game has three Nash equilibria, {Up, $y = 100$}, {Right, $x = y = 0$}, and a mixed one where A mixes Up with probability 3/5 and Right with probability 2/5, and B mixes $x = 0, y = 100$ with probability 1/3 and $x = y = 0$ with probability 2/3.

As mentioned earlier, central elements of impulse learning models are the confidence with which a player chooses a pure strategy k and the impulse caused by the feedback players receive on their previous choice of k. Typically, impulse learning models interpret the confidence for k as the density of k (i.e., k's attraction) to be chosen in the consecutive period. For our analysis, we will offer a different reading of confidence: we assume that B's increasing confidence for k corresponds with an increasing number of points B assigns for k. That is, players assign points in accordance to their confidence. Similar behavior is documented elsewhere and referred to as probability matching (e.g., Vulkan, 2000). In turn, changes in the points assigned on k show us the effect of feedback on the confidence for k – in other words, the impulse.

3 Experimental Procedure and Expectations

The game is repeated twice with constant roles, but changing partners. We have two treatment conditions in the experiment: a game with no feedback

(NF),[3] and a game with partial feedback (PF). In the NF treatment, neither player A nor player B receives any feedback on payoffs or the opponent's decision until the end of the experiment; NF serves as a baseline. In the PF treatment, player A receives no feedback like in NF, while player B is informed about the δ_A in her first game before the second starts. Common knowledge are the constant roles, the feedback setting, the random rematching with a new partners, and that one of the two games is randomly drawn at the end of the experiment and paid out.

The experiments were run in the WisoLab at the University of Hamburg in mid 2011. In the experiment, we use a graphical representation of (1) and (2) (see the experimental instructions in the Appendix). Printed instructions were distributed among the participants (in total, 136 students from various disciplines at the University of Hamburg, 53 percent females, median age 23). The experiment was conducted with the software Z-Tree (Fischbacher, 2007); the participants were recruited with ORSEE (Greiner, 2004). Each session lasted about 30 minutes, average earnings were 6.10 Euros, plus a show-up fee of 4 Euros.

Before focusing on player B's learning, we have to distinguish in a first step between "good news" and "bad news" for B's strategy. Therefore, we have to define a reference value that divides the return rates per point into superior and inferior outcomes. Following Selten and Chmura (2008), we use the maximum of B's minimum values ("maximin value") for this purpose. That is, we choose the return rate B receives for sure as the reference value. In our game, any point distributed on x yields at least one point (i.e., $\frac{\partial \pi_B}{\partial x} \geq 1$), while the other two alternatives may lead to lower returns. Therefore, B's maximin value is one, while return rates smaller (greater) than one are denoted as inferior (superior). Consequently, "good news" for x are $\delta_A =$ Left, for y $\delta_A =$ Up, while "bad news" for y are $\delta_A =$ Right.[4] It follows for positive impulse learning models (e.g., reinforcement learning like Erev & Roth, 1998, or experience weighted attraction learning, like Ho et al., 2008) that feedback of $\delta_A =$ Left ($\delta_A =$ Up) in the first period increases the number of points assigned on x (y) in the second period, while negative impulse-learning models (e.g., impulse matching learning like Chmura et al., 2012) assume that feedback of $\delta_A =$ Right in the first period decreases the number

[3]Notice that Weber (2003) and Scott & Weber (2010) find systematic learning in a game without feedback.

[4]The number of points assigned to $100 - x - y$ is special in the sense that it is by itself a residual rather than a clear-cut option. Therefore, subjects may not learn with respect to this alternative as they learn predominantly on x and y. Nonetheless, one can consider "good news" ("bad news") for $100 - x - y$ as $\delta_A =$ Right ($\delta_A =$ Up and $\delta_A =$ Left). We will acknowledge the special character of this alternative in the consecutive analysis.

of points assigned on y in the second period. Notice that no impulse model provides specific predictions concerning the alternative which loses points (to counterbalance a positive impulse), or gains points (to counterbalance a negative impulse).

4 Results

Let us first consider B's decisions in the first game. On average, B invests approximately 50 points altogether on x and y, both in NF and PF. Mean x is 31.1 (25.9) and mean y is 26.7 in NF (30.3 in PF); the differences between NF and PF are not significant (Wilcoxon-Rank-Sum-Test, two-tailed, $p = 0.33$ for x, and $p = 0.59$ for y). Moreover, x and y are significantly positive correlated both in NF and PF.[5] Thus it seems that Bs divide points roughly equally between x and y.

The majority of As favors Right (59% in NF, 56% in PF), while only few choose Up (18% in NF, 23% in PF) or Left (23% in NF, 21% in PF) in the first period in both treatments (again, no statistical difference between the treatments: $p = 0.83$, chi-sqare test, two tailed). This pattern is similar in the second period. Yet, although the difference between the treatments is insignificant ($p = 0.26$, chi-square test, two tailed), it seems that As choose more consistently in the second and the first period of PF (65% Right, 12% Up, 23% Left) than in NF (82% Right, 6% Up, 12% Left). As may want to facilitate coordination once their decisions are observed, despite the rematching between games.

Since B's decisions on x and y are interdependent, we analyze the changes of points on x and y between the first and second period by means of a simultaneous equation estimation. The dependent variables x and y in the second period, denoted as x_2 and y_2, are regressed on x_1 and y_1, respectively (i.e., x and y in the first period), in order to test for the path dependency of learning.[6] In addition, we test the impulse of feedback on x_2 and y_2: For this purpose, let us define the dummy variables f_u, f_l and f_r which are one if B receives in PF the feedback that $\delta_A =$ Up, $\delta_A =$ Left, and $\delta_A =$ Right the first period, respectively, and zero otherwise. Regression results are shown in Table 1.[7]

[5] Correlations are 0.39 in NF and 0.39 in PF, $p = 0.02$ and $p = 0.02$, two-tailed correlation test.

[6] One could claim that we do not observe some kind of learning, but simply hedging between risky actions. In this case, we should observe significantly negative coefficients x_1 and y_1.

[7] Standard errors in parentheses; asterisks indicate levels of significance: * significant at a 10% level, ** significant at a 5% level, and *** significant at a 1% level; number of

Table 1: *Estimated coefficients for the impulse on x_2 and y_2*

independent	dependent	
	x_2	y_2
x_1	.789*** (.126)	
y_1		.258** (.131)
f_u	−12.31 (10.99)	7.04 (11.83)
f_l	−1.28 (7.18)	−.51 (7.46)
f_r	10.09* (6.08)	−13.34** (6.48)
constant	9.17* (5.44)	26.67*** (5.29)
nob	68	68
r-square	.34	.09
chi-square-test	46.13***	8.94*

Three results of the estimation are remarkable: first, the significant positive coefficients of x_1 and y_1 show the path-dependency of B's decision. Thus, there is evidence for gradual adjustments of points (i.e., learning) even in a two-period game and still without feedback. Second, the significant negative coefficient of f_r in the equation on y_2 suggests that impulse learning based on negative feedback takes place, while there is no evidence for impulse learning based on positive feedback.[8] Finally, there is at least weak evidence (i.e., the weakly significant coefficient of f_r in the equation on x_2) that the alternative x, ensuring at least the reference value, counterbalances the negative impulse on y. This is remarkable since increasing x in not a (myopic) best response to δ_A =Right in the first game.

5 Discussion

Let us summarize our results: First, we are able to show that the impulse is triggered predominantly by "bad news". That is, players who receive a payoff below a certain reference value by choosing action k decrease the confidence with which k is selected in the consecutive period. In the meantime, the confidence for the alternative action which yields the reference value for sure increases in a weakly significant way. While we cannot generalize our results

independent observations (*nob*), the r-square, and a chi-square-test for the goodness-of-fit are reported.

[8]Notice that there is no evidence on positive nor negative feedback learning for $100 - x - y$. This, however, may result from its special character (see footnote 4).

to the scenario of multiple repeated interactions (perhaps impulses based on "good news" require several stimuli), it seems that learning at least in short sequences of repeated play is facilitated by negative feedback rather than positive feedback.

References

[1] Beggs, Alan W. (2005), On the convergence of reinforcement learning, *Journal of Economic Theory* **122**, 1–36.

[2] Camerer, Colin F. & Teck H. Ho (1999), Experience-weighted attraction learning in normal form games, *Econometrica* **67**, 837–874.

[3] Chmura, Thorsten, Sebastian J. Goerg & Reinhard Selten (2012), Learning in experimental 2 x 2 games, *Games and Economic Behavior* **76**, 44-73.

[4] Erev, Ido & Alvin E. Roth (1998), Prediction how people play games: Reinforcement learning in games with unique strategy equilibrium, *American Economic Review* **88**, 848–881.

[5] Erev, Ido & Ernan Haruvy (2012), Learning and the economics of small decision. Forthcoming in: John H. Kagel and Alvin E. Roth (eds.), *The Handbook of Experimental Economics*, 2nd Edition, Princeton University Press.

[6] Fischbacher, Urs (2007), Z-Tree: Zurich toolbox for ready-made economic experiments, *Experimental Economics* **10**, 171–178.

[7] Greiner, Ben (2004), An online recruiting system for economic experiments. In: Kurt Kremer and Volker Macho (eds.), *Forschung und wissenschaftliches Rechnen 2003*, Göttingen, 79–93.

[8] Grosskopf, Brit (2003), Reinforcement and directional learning in the ultimatum game with responder competition, *Experimental Economics* **6**, 141–158.

[9] Ho, Teck H., Colin F. Camerer & Juin-Kuan Chong (2008), Self-tuning experience weighted attraction learning in games, *Journal of Economic Theory* **133**, 177-198.

[10] Marchiori, Davide & Massimo Warglien (2008), Predicting human behavior by regret driven neural networks, *Science* **319**, 1111–1113.

[11] Ockenfels, Axel & Reinhard Selten (2005), Impulse balance equilibrium and feedback in first price auctions, *Games and Economic Behavior* **51**, 155–170.

[12] Rick, Scott & Roberto A. Weber (2010), Meaningful learning and transfer of learning in games played repeatedly without feedback, *Games and Economic Behavior* **68**, 716–730.

[13] Roth, Alvin E. & Ido Erev (1995), Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term, *Games and Economic Behavior* **8**, 164–212.

[14] Selten, Reinhard, Klaus Abbink & Ricarda Cox (2005), Learning direction theory and the winner's curse, *Experimental Economics* **8**, 5–20.

[15] Selten, Reinhard & Thorsten Chmura (2008), Stationary concepts for experimental 2x2-games, *American Economic Review* **98**, 938–966.

[16] Vulkan, Nir (2000), An economist's perspective on probability matching, *Journal of Economic Surveys* **14**, 101–118.

[17] Weber, Roberto A. (2003), 'Learning' with no feedback in a competitive guessing game, *Games and Economic Behavior* **44**, 134–144.

Appendix: Instructions[9]
– not intended for publication –

General Rules

Welcome to the experiment. Please read the following instructions carefully. It is very important that you do not talk to other participants during the experiment. If you have questions about the experiment, please contact us. One of the experimenters will come to you and clarify the issues.

In the experiment you are asked tointeract anonymously with the other participants of this experiment. All participants received identical instructions. At no time can any other participant link your decisions with your identity.

[9]The following instructions are translations of the German originals for the NF condition. Variations in the PF condition are indicated by brackets.

The game

In this experiment, you will play the following game twice. To distinguish the games, we call the first game the "blue game" and the second game the "green game". In each of the two games, there are two roles that are player A and player B. At the beginning of the experiment, you will be randomly assigned a role. In both games, you will hold the same role. In each game, you are randomly and anonymously matched with one player of the other role. Player A will be matched with player B. Player B will be matched with player A. You never interact with the same player twice. That is, for each of the two games you are assigned a new partner. For each of the two games you only get feedback about the behavior of the other party after you have decided for both games. [After player B has decided in the blue game, she learns how player A behaved in a blue game. Player A receives feedback about the behavior of the other party only after she has decided for both games.]

During the experiment, we speak of points. Your total income will be calculated first in points. At the end of the experiment, all the points you earn in the experiment will be converted into Euros with the following exchange rate:

$$20 \text{ points} = 1 \text{ Euro}$$

At the end of the experiment, one game will be randomly drawn from the two. Only the payment of the drawn game is payoff-relevant, i.e., only the points you earn in the selected game determine your earnings. In addition, you will receive 4 Euros for your participation regardless of your behavior. You will receive the total amount from us in cash. On the next page, we explain the special rules of this experiment.

The game has two stages:

Stage 1: The player in the role of A chooses Up, Left, or Right.

Stage 2: The player in the role B receives 100 points. Without knowing the choice of player A, player B must now decide how many points x out of the 100 points she puts in the first account, and how many points y out of the 100 points she puts in the second account (x and y are both numbers between 0 and 100; the sum of x and y cannot exceed 100). B retains the rest, that is, $100 - x - y$.

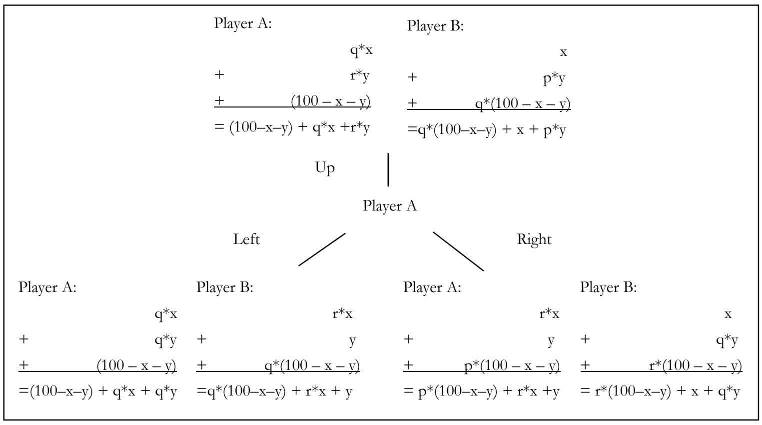
The payoffs of the players are calculated as follows:

If player A chooses Up, player A earns $(100 - x - y) + qx + ry$ points and player B earns $q(100 - x - y) + x + py$ points.

If player A chooses Left, player A earns $(100 - x - y) + qx + qy$ points and player B earns $q(100 - x - y) + rx + y$ points.

If player A chooses Right, player A earns $p(100 - x - y) + rx + y$ points and player B earns $r(100 - x - y) + x + qy$ points.

In the actual games, you will see numbers instead of letters p, q, and r. The numbers remain the same in the blue and green games, i.e., the blue and the green games are identical. You will play the game twice with different people, but you will always play the same role. Please choose first for the blue game and then for the green game. The following figure shows the game again schematically:



Your payment

At the end of the experiment, one game will be randomly chosen as your final payoff.

If you have any questions, please contact us now.