Distracting from Equilibrium: How Feedback Can Reinforce Heuristic Play in Strategic Games

Sibilla Di Guida[a,*], Davide Marchiori[a], Damien Mayaux[a,b], Luca Polonio[c]

[a] *IMT School for Advanced Studies Lucca, Piazza S. Ponziano, 6 - 55100 Lucca, LU, Italy*
[b] *Paris School of Economics, 48 Boulevard Jourdan, 75014 Paris, France*
[c] *University of Milan - Bicocca, Piazza dell'Ateneo Nuovo, 1, 20126 Milano MI, Italy*

Abstract

In this study, we examine the hypothesis that, when playing a sequence of games, players may use feedback information to reinforce heuristics of play instead of learning equilibrium play, even when such information is immediate and accurate. This departs from the general view that feedback in repeated interactions eventually leads to equilibrium play. To test this, we designed an experiment with participants facing sequences of 2-person 3x3 games incorporating five common heuristics of play. Our 5x2 factorial design includes two learning settings for each heuristic: "overlapping strategies" where the target heuristic and equilibrium strategy lead to the same action, and "distinct strategies" where they differ. This setup allows comparison of learning outcomes across target heuristics. Results show that learning equilibrium play is more challenging in overlapping strategies treatments, with many players interpreting feedback as reinforcing the target heuristic. This effect varies across heuristics. Our findings demonstrate that even complete, unambiguous feedback can reinforce non-rational behaviors. We conclude that caution is needed when interpreting observed choice behavior in repeated strategic interactions. Observed equilibrium choices may not reflect rational play, and accurate feedback can reinforce non-equilibrium strategies. We discuss methodological implications for the field.

Keywords: Strategic thinking, heuristics, learning, experiment, equilibrium play, focal point
JEL: C72, C73, C91

*Corresponding author: sibilla.diguida@imtlucca.it

1. Introduction

In commonly studied strategic interactions, feedback information is typically leveraged to facilitate learning of the rational course of action (Fudenberg and Levine, 1998, 2009; Erev and Haruvy, 2016). Whereas learning can also happen without any information about the decision outcomes (Selten and Chmura, 2008; Weber, 2003), the observed effects on choice behavior are relatively smaller than when feedback is provided. Behavioral game theory studies show that repetition and feedback generally guide behavior towards equilibrium play (Hertwig and Ortmann, 2001; Camerer et al., 2004). Quick convergence to equilibrium is especially observed in simultaneous-move, 2x2 matrix games, when feedback is complete, accurate, and immediate. In these settings, players are informed about their and their counterparts' choices and outcomes, this information is certain and unambiguous and is provided right after each choice. These are the feedback conditions we examine in our study.

Since learning is commonly defined as the effect of experience on choice behavior, most learning studies use the average frequency of equilibrium choices over time as a unit of analysis (learning trajectories - see Erev and Roth (1998); Ho et al. (2007); Marchiori and Warglien (2008)). This measure, however, does not reflect what players actually learn; an increase in the rate of equilibrium choices does not necessarily indicate an increased intent to play rationally, as other intents may cause it. Marchiori et al. (2021) provides one of the first attempts to clarify this issue, showing that games of different strategic/cognitive complexity (dominance vs. iterated dominance games) offer different learning opportunities, even with feedback. Results show that whereas the frequency of equilibrium choices increases over time in both game types, eye-tracking data reveals that the majority of participants adopt equilibrium-compatible information gathering only in iterated-dominance games. Salmon (2001) pioneering research showed the complexity and difficulty of trying to accurately and reliably identify which learning rules are used by subjects in experimental settings. His results indicate that even the most common econometric models have difficulty in accomplishing such task. Although these studies establish that feedback is not a sufficient condition for the learning of rational play (beyond a mere increase of the rate of equilibrium choices), they do not explore what heuristics players may learn, or how these heuristics may hinder learning of equilibrium thinking.

Our study addresses this gap by systematically: 1) analyzing players' choice behavior when learning occurs in games where the equilibrium choice differs from the choice suggested by a target heuristic, compared with games where these choices overlap. This comparison allows us to quantify the extent to which target heuristics interfere with equilibrium learning. 2) Comparing the "distraction from equilibrium" effect across five commonly observed target heuristics. This comparison reveals how learning outcomes vary across the different heuristics.

Our results have significant theoretical and methodological implications. Theoretically, we show that feedback is not sufficient for correcting players' beliefs and controlling their interpretation of the game situation. Surprisingly, even complete, unambiguous feedback can reinforce non-rational behaviors. From a methodological perspective, our findings urge caution in interpreting and generalizing data from learning experiments. We warn against equating an increased frequency of equilibrium choices (observable) with actual learning (generally non-observable). Although learning can be inferred from experiments carefully designed for this purpose, it is crucial to recognize that observed behavior may not directly reflect the underlying learning process.

2. Theoretical framework and hypotheses

In behavioral game theory, learning commonly indicates "an observed change in behavior owing to experience" (Camerer, 2011, p.265), and an increase in the frequency of equilibrium choices is often implicitly equated to a correspondingly increased level of strategic sophistication. This leads to the popular conception that players choose equilibrium actions because they have (at least to some extent) learned to analyze a game rationally. As Camerer (2011, p.265), puts it, "Equilibrium concepts implicitly assume that players either figure out what equilibrium to play by reasoning, follow the recommendation of a fictional outside arbiter... or learn or evolve toward the equilibrium". We challenge this interpretation, emphasizing the need to distinguish between the observed adaptation of choice behavior and its inferred qualitative content. Specifically, we show that what players learn with feedback depends upon the structure of a game, and that equilibrium play may mask underlying non-rational intents. This because players may use feedback to reinforce choice strategies that are only incidentally consistent with equilibrium play. We show that this phenomenon occurs when a game's

	C1	C2	
R1	1 1	**3** **3**	Equilibrium, Focal Point
R2	2 4	4 2	

Notes: Within each cell, the bottom left (top right) value is the payoff of the row (column) player. The top right cell, with its high and symmetrical payoffs, is a focal point (underlined). The top right cell is also the unique Nash equilibrium in pure strategies (in bold).

Table 1: A game in which the Focal Point and Equilibrium actions overlap

equilibrium strategy matches the one suggested by an alternative heuristic of play.

Consider the scenario of a player (Row Player) facing the game presented in Table 1, competing against a rational, profit-maximizing algorithm (Column Player). The algorithm will select C2, its dominant strategy.

If feedback is provided, all of the player's strategies that lead to the selection of R1 will be reinforced. Equilibrium play will be reinforced, but also trying to coordinate on cell (R1, C2). This latter action, which maximizes Social Welfare and minimizes Social Distance, is a natural *focal point* - see (Schelling, 1960; Cooper et al., 1990; Van Lange, 1999, 2000; Crawford et al., 2008; Jackson and Xing, 2014; Parravano and Poulsen, 2015; Polonio et al., 2015; Devetag et al., 2016; Polonio and Coricelli, 2019). As feedback reinforces both strategies, and the player's best response to each is R1, solely observing the chosen action, we cannot infer the player's actual underlying intent. However, observing choice behavior in a second, opportunely designed game can help disentangle between the two possibilities. Suppose that after playing games similar to that in Table 1, a player faces games like that in Table 2. Here, different strategic intentions lead to distinct choices. Players consistently adopting a rational strategy, would now select R1. Conversely, those who try to coordinate on the focal point would choose R2.

Ambiguous situations like this can arise when the same piece of feedback information reinforces different strategies. We argue that when multiple strategies overlap on the same action, different interpretations of feedback can be equally plausible. Specifically, some players will see it as reinforcing equilibrium play, others as reinforcing alternative heuristics of play. This

	C1	C2	
R1	1 1	**2** **2**	Equilibrium
R2	<u>3</u> <u>3</u>	4 1	Focal Point

Notes: The top right cell is the unique Nash equilibrium in pure strategies (in bold). The bottom left cell, with its high and symmetrical payoffs, is a focal point (underlined).

Table 2: A game in which the Focal Point and Equilibrium actions are distinct

leads to Hypothesis 1:

Hypothesis 1. *In ambiguous*[1] *strategic settings, we expect some players to interpret feedback information as reinforcing the rational course of action. However, a significant share of players will interpret it as reinforcing some alternative heuristic of play.*

Hypothesis 1 is particularly significant from a theoretical perspective. Our study suggests that the design of a game itself can influence how players use feedback information, and that complete and unambiguous feedback can lead to different learning outcomes.

The elimination (or significant reduction) of strategic ambiguity facilitates players' identification of a game's equilibrium structure. This leads to the following Hypothesis 2:

Hypothesis 2. *In unambiguous strategic situations, players are expected to identify the (pure) equilibrium strategy more readily than in ambiguous settings. Consequently, we anticipate that players will learn equilibrium play more frequently in unambiguous settings than in ambiguous ones.*

Our hypotheses so far hinge on the structural ambiguity of a game, as previously defined. However, the degree of structural ambiguity is expected to vary depending on which heuristic aligns with the equilibrium strategy. We anticipate that the more salient, widely diffused across the population,

[1]In our study, the adjective "ambiguous" always refers to the concurrent existence of different strategic intentions within the same action, not to uncertainties in the features of the game structure.

and easily recognizable a heuristic is, the stronger its distraction effect from equilibrium play will be. It is important to note that the plausibility and prevalence of a heuristic in the population may result from a selection process based on its effectiveness among a population of boundedly rational players. On this point, the study by Spiliopoulos and Hertwig (2020) links the prevalence of a heuristic in the population to its performance effectiveness. They demonstrate, for example, that the L1 heuristic (which we also consider in our study; see next section for a description) is the most widespread as it yields the highest average payoff compared to other heuristics.

Let us illustrate this with an example. In the game in Table 1, the equilibrium strategy aligns with what is usually referred to as the *focal point strategy*. Research has shown that people frequently use focal points as coordination devices (Schelling, 1960; Mehta et al., 1994; Parravano and Poulsen, 2015), and symmetric game outcomes with high rewards for both players are generally perceived as focal points (Goeree and Holt, 2001; Crawford et al., 2008; Jackson and Xing, 2014). Thus, the focal point strategy consists of selecting the option that includes the focal outcome. In these games, it is reasonable to expect a substantial proportion of players to learn a "focal-point seeking" strategy rather than the equilibrium one.

Now consider a different game, presented in Table 3. Here, the algorithm will select C2, its dominant strategy. Players rationally best responding to this strategy will choose R1, as would players adopting the Optimistic heuristic (i.e., aiming for their highest possible payoff). In this case, feedback information would reinforce both rational play and the Optimistic heuristic. However, since the Optimistic heuristic is not much diffused among players (Selten et al. (2003)), learning from these games is likely to enhance rational strategic thinking in a comparatively larger share of the population.

In conclusion, comparing behavior in the games presented in Table 2 and 3 is expected to show a higher proportion of heuristic-consistent choices in the former than in the latter. This expectation forms the basis for our Hypothesis 3:

Hypothesis 3. *Heuristics differ in their plausibility. Thus, in strategically ambiguous situations, the frequency of heuristic-consistent choices will depend on the specific heuristic reinforced by feedback information.*

We will test this hypothesis by considering learning outcomes across five target heuristics, selected among the most studied ones, and differing in their inherent features.

	C1	C2	
R1	1 1	**2** **4**	Equilibrium, Optimistic
R2	2 3	3 1	

Notes: The top right cell is the unique Nash equilibrium in pure strategies (in bold). It also corresponds to the highest possible payoff for the row player (underlined).

Table 3: A game in which the Optimistic and Equilibrium actions overlap

2.1. Selected heuristics

One of the goals of this paper is to test how different heuristics influence learning outcomes. To this aim, we selected five heuristics that are among the most studied in the literature (Costa-Gomes et al., 2001; Costa-Gomes and Weizsäcker, 2008; Spiliopoulos and Hertwig, 2020). Of these five heuristics, two are widely adopted (other regarding Focal Point and Level 1), two are less frequently observed (Optimistic and Level 2), and one is particularly relevant in our specific context, as we will explain later (Best Reply to an Optimistic counterpart). Among these heuristics, three are defined as strategic (Focal Point and Level 2), whereas the others are not (Level 1 and Optimistic). Table 4 summarizes the characteristics of these heuristics.

As detailed in section 3, players in our experiments are matched with a profit-maximizing algorithm. Such a design ensures that all participants make decisions within the same strategic environment. To communicate the algorithm's strategic behavior in both precise and accessible terms, players were informed that the algorithm "will try to earn as much as possible, will assume you will do the same, and will not change strategy through the experiment." The five heuristics we selected align to varying degrees with this description of the algorithm's behavior. Therefore, the plausibility of these heuristics is also expected to vary based on this information.

Focal Point: When considering a course of action, agents might analyze a game focusing on outcomes, rather than by strategy. In such cases and based on individual prosocial attitudes, analyses might include looking for outcomes that maximize Social Welfare (the sum of the player's own and the opponent's payoffs), or minimize Social Distance (the difference between the player's own and the opponent's payoffs) (Van Lange, 1999, 2000; Jackson and Xing,

Strategy	Name	Considers the other's payoff	Outcome-Based	Level of complexity	Belief consistent with instructions	Description
Level-1	L1	—	—	1	—	Choose the action(s) best responding to the assumption that an opponent is choosing randomly
Level-2	L2	✓	—	2	✓*	Choose the action(s) best responding to the assumption that an opponent is applying L1
Optimistic	OPT	—	✓	1	—	Choose the action(s) offering the highest payoff for the player
BR to Optimistic	BRO	✓	—	2	✓*	Choose the action(s) best responding to the assumption that an opponent is applying OPT
Focal Point	FP	✓	✓	3	✓	Choose the action(s) maximizing the sum and minimizing the difference of the player's and the opponent's payoff
Equilibrium	EQ	✓	—	3	✓	Choose the action(s) consistent with the Nash equilibrium

Notes: The "level of complexity" column should be read as follows: 1) players focus solely on their own payoffs; 2) players assume that their opponent focus solely on its own payoffs; 3) players assume the opponent takes into account also the players' payoffs when choosing its action. The "Belief consistent with instructions" column should be read as follows: - indicates no consistency; ✓* indicates a naive consistency, where the instructions provided are only partially understood; ✓ indicates a full consistency.

Table 4: Description of the target heuristics and their characteristics

2014; Polonio and Coricelli, 2019; Spiliopoulos and Hertwig, 2020)[2]. In each game, we create a Focal Point (FP) as an outcome that maximizes Social Welfare while maintaining a Social Distance of 0 (with symmetric payoffs). This suggests that when reinforced, it could emerge as a highly salient and efficient strategy. In our experimental design, feedback that supports the belief that the algorithm might attempt to coordinate on the game's FP aligns with the described algorithm behavior, making this heuristic highly plausible and salient in our settings.

Level 1: The Level 1 heuristic (L1) often represents a reasonable and cognitively inexpensive approach to play (Spiliopoulos and Hertwig, 2020). Evidence shows that players using this non-strategic heuristic focus on their own incentives, disregarding those of their counterpart, and select the option with the highest average payoff (Devetag et al., 2016). Within our experimental framework, the assumption that L1 players assume their counterpart to choose randomly is inconsistent with the information provided in the instructions. Nonetheless, as long as feedback reinforces this strategy, players might lack the motivation to explore more complex strategies that would require greater cognitive effort.

Level 2: The Level 2 heuristic (L2) involves more sophisticated strategic thinking than L1. Players using L2 are assumed to anticipate that their opponent will behave as a L1 player and will then choose the best response to the expected move (Nagel, 1995; Stahl and Wilson, 1995). This two-step reasoning categorizes L2 as a strategic heuristic. Although the L2 heuristic, in 2x2 games, invariably corresponds to a pure equilibrium strategy if there exists one, this does not hold for 3x3 games. Thus, to be able to keep the L2 choice separate from the equilibrium choice, we only considered 3x3 games in our experimental design (see Section 3). Although attributing to the counterpart an L1 behavior would not fit the description provided, applying an L2 strategy could be motivated by a naive interpretation of the information once supported by reinforcing feedback.

Optimistic: Players using the Optimistic heuristic (OPT) target the option with the highest possible payoff, disregarding their opponent's payoffs. This non-strategic heuristic represents strategically naive behavior, which is

[2]Of course, outcome-based strategies might also include heuristic based on negative prosociality. However, as discussed at the end of this section, such heuristics are not relevant in our context.

uncommon in interactive settings (Spiliopoulos and Hertwig, 2020). Applying the Optimistic strategy does not align with the provided information about the algorithm's strategy, suggesting that agents using this heuristic disregard the given information. Consequently, we expect that positive reinforcement will have only a minimal impact on the frequency and salience of Optimistic behavior.

Best Reply to Optimistic: We included the "Best Reply to Optimistic" (BRO) heuristic. Although mostly unexplored in the literature, this heuristic is relevant to our context, as it aligns with a possible (albeit imprecise and naive) interpretation of the description of the algorithm's strategy provided to players. This heuristic assumes that players respond optimally to an Optimistic player who targets their highest payoff in the game matrix. Thus, a naive reading of experimental instructions may suggest to some players that the algorithm selects actions trying to obtain its largest possible payoff, ignoring the strategic dimension of the game interaction. Such interpretation would be supported by feedback.

Notably, we focused on heuristics pertinent to our experimental setting of human-algorithm interaction, omitting those that are not. We collapsed heuristics embedding positive social preferences such as maximizing joint payoff (Costa-Gomes et al., 2001) and minimizing payoff distance (Fehr and Schmidt, 1999) within the Focal Point heuristic. Conversely, we omitted heuristics embedding negative social preferences, such as heuristics where players attempt to maximize the difference between their payoff and the opponent's. These are inconsistent with the provided information about the algorithm's strategy. Our reasoning for this exclusion and simplification is twofold: 1) The design constraints of 3x3 games (see next section for the details) limit the number of distinct heuristics that can be implemented without overlap; including competitive heuristics would have inevitably led to redundancies with those already in place. 2) Parsimony and feasibility considerations; incorporating additional heuristics would have significantly expanded the number of treatments, compromising the study's manageability and focus. Finally, it is important to highlight that in terms of game design, we designed games with a unique pure strategy Nash Equilibrium, and avoided games with dominance, as the presence of a dominant strategy (a powerful attractor of choice behavior) would have significantly diminished the effect of the heuristics under study.

3. Experimental design

The experiment consists of several treatments composed of three sequential stages: Assessment, Learning, and Reassessment. A similar design was adopted in Marchiori et al. (2021). In the initial Assessment stage, participants play 14 2-person, 3×3 games without receiving feedback about their counterpart's choices and the obtained payoffs. In this stage, participants' initial strategic skills are assessed. In the subsequent Learning stage, participants play 14 games (details provided later in this section) receiving feedback after each choice. Feedback information summarizes both the actions chosen by the algorithm and those of the player, indicating their payoffs. By varying the kind of games that are presented in the Learning stage (which differ across treatments), we can control for participants' accumulated experience, and check its effects on the learning outcome. The final Reassessment stage reassesses participants' strategic skills after learning has shaped them: participants play another sequence of 14 games structurally similar to those played in the Assessment, again without receiving any feedback. Such a design allows us to evaluate the effects of the intermediate Learning stage by comparing, within-subject, choice behavior in the first and final stages.

To enhance control over players' beliefs about their counterpart's behavior, players were matched against an algorithm. Matching players against each other would have raised expectations and shaped beliefs about their counterparts that would have been difficult to control for. Players were all equally instructed that they would play against an algorithm that would choose to maximize its payoff, under the assumption that its opponent (i.e., the human player) would do the same. They were also informed that the algorithm would not adjust its strategy based on previous interactions, remaining consistent throughout the experiment.

To examine whether feedback in ambiguous settings affects participants differently from feedback in unambiguous ones, we employed a 2x5 factorial design, with two levels of ambiguity (overlapping/distinct strategies) for each target heuristic - see Table 5. Within each target heuristic, the treatments differ only for the games in the Learning stage when feedback is provided to players. In this stage, the Overlapping treatment composes entirely of games in which the target heuristic strategy and the equilibrium strategy are the same (ambiguous setting), while the Distinct treatment includes games in which the target heuristic strategy and the equilibrium strategy are distinct (unambiguous setting). Moreover, while the games in stages 1 and 3 are

	Level of ambiguity			
	Overlapping L1	(N=83)	Distinct L1	(N=83)
Target	Overlapping L2	(N=95)	Distinct L2	(N=78)
heuristic	Overlapping OPT	(N=76)	Distinct OPT	(N=88)
	Overlapping BRO	(N=80)	Distinct BRO	(N=77)
	Overlapping FP	(N=94)	Distinct FP	(N=73)

Table 5: The 2×5 experimental treatments and their respective number of subjects

identical between the two treatments, the Learning stage includes games that are similar (see Section 3.1 for a detailed explanation). A between-subjects analysis of behavior in stage 3 across the two treatments enables us to assess how choice behavior has been influenced during the learning stage.

As stated in Hypothesis 3, the impact of learning in ambiguous settings is expected to differ across target heuristics. To test this hypothesis, we applied the same learning design using ambiguous vs. unambiguous games for each of our five target heuristics, yielding ten experimental treatments, as outlined in Table 7.

3.1. Game structure

The game matrices for this experiment were carefully designed to allow for the comparison of results across treatments. Each game was designed to embed all five heuristics (L1, L2, FP, OPT, and BRO), plus the (unique) pure strategy Nash equilibrium (EQ). For concreteness, we discuss the rationale for the game design considering the target heuristic Level 1, with the same approach applying to all other heuristics.

The Level-1 treatments. These two experimental treatments consider the Level 1 heuristic as the target. As mentioned earlier, the two experimental treatments (labeled Overlapping-L1 and Distinct-L1), differ on whether the strategy for the target heuristic is identical to the equilibrium strategy, in the games presented in the Learning stage. In both treatments, the strategy for heuristics FP, L2, and OPT is associated with Row 3, whereas the BRO strategy is associated with Row 2[3]. In the Overlapping treatment games,

[3]In Subsection 3.2 it is explained why BRO is associated with Row 2 and not Row 3, as the other heuristics are.

	C1	C2	C3	
R1	5.5 *6.5*	5 6	**6.1** **7.1**	Equilibrium, L1
R2	6.6 *5.1*	4.5 7.2	7 6.3	BRO
R3	8 8	9 4	5.3 5.6	L2, OPT, FP

Overlapping-type game

	C1	C2	C3	
R1	5.5 *5.1*	5 6	**6.1** **7.1**	Equilibrium
R2	6.6 *6.5*	4.5 7.2	7 6.3	BRO, L1
R3	8 8	9 4	5.3 5.6	L2, OPT, FP

Distinct-type game

Notes: Within each cell, the bottom left (top right) value is the payoff of the row (column) player. The labels to the right mark the row corresponding to each heuristic. In the first game, the top-row action is consistent with both Equilibrium and the L1 heuristic. In the second game, the top-row (middle-row) action is consistent with Equilibrium (L1 heuristic). The two matrices differ only for the row player's payoff in the top and middle left cells. The Equilibrium choice is in bold, whereas the payoffs that differ between the two games are in italics.

Table 6: Example of Overlapping- and Distinct-type games for the L1 heuristic

L1 and Nash players would choose the same action, Row 1, whereas in the Distinct treatment games, Nash players would choose Row 1 and L1 players Row 2. Table 6 represents our baseline Overlapping and Distinct games for the L1 heuristic. From these baseline matrices, the 28 matrices used for the experiment were created through random transformations of the payoffs that did not alter the described structure.

In the Assessment and Reassessment stages, participants play a sequence of games including seven Overlapping and seven Distinct games (not identical to those proposed in the Learning stage), without receiving feedback. Conversely, in the Learning stage, agents play 14 games of the same type

(only Overlapping or only Distinct), with feedback. With such a game structure, in treatment Distinct-L1, feedback indicates that the best strategy is the equilibrium choice. In treatment Overlapping-L1, players may learn to identify the equilibrium strategy or the L1 heuristic one, depending on their interpretation of feedback.

Level-2, Focal Point, Optimistic, and Best Reply treatments. Similarly to what we discussed for the Level 1 treatments, we created different game matrices for each of the other treatments: Level 2, BRO, Focal Point, and Optimistic, for a total of 10 baseline matrices ({Heuristics} × {Overlapping, Distinct}). The main scheme was the same: in the Distinct games, the equilibrium strategy was assigned to Row 1, and the target heuristic to Row 2; conversely, in the Overlapping games, both the target heuristic and the equilibrium strategy were associated with Row 1. All other heuristics were assigned to Row 3, except for the BRO heuristic assigned to Row 2. However, when the BRO heuristic served as the target, it was placed on Row 2 in the Distinct treatment and on Row 1 in the Overlapping treatment (all other heuristics being in Row 3). As done for the L1 treatments, we designed the baseline matrices and generated different instances of them by applying random perturbations to payoffs that did not alter the original structure. All baseline matrices are presented in Appendix G. Table 7 provides a summary of the experimental treatments, while Table G.20 shows, for each treatment, which row of game matrix each heuristic was on.

3.2. Game construction rules

The games presented in the experiment were carefully designed to satisfy some general rules in addition to the rationale exposed earlier. These rules were established to minimize potential confounding factors. All games share the following characteristics:

- There is a unique Nash equilibrium in pure strategies.
- The Level-1 strategy yields an average payoff that is notably higher than the one of the other two rows.[4]

[4]The average payoff of the L1 row is on average 5.34% higher (min 0.5%) than the second highest and on average 8.57% higher (min 2.3%) than the mean of the two other rows. The same holds for the column player, where the average payoff of the L1 column is on average 8.3% higher (min 1.1%) than the second highest and on average 14.2% (min 3.5%) higher than the mean of the two other columns

Treatment	Stage 1: Assessment	Stage 2: Learning	Stage 3: Reassessment
Overlapping L1	7 Overlapping games for L1 7 Distinct games for L1	14 Overlapping games for L1 Feedback	7 Overlapping games for L1 7 Distinct games for L1
Distinct L1	No Feedback	14 Distinct games for L1 Feedback	No Feedback
Overlapping L2	7 Overlapping games for L2 7 Distinct games for L2	14 Overlapping games for L2 Feedback	7 Overlapping games for L2 7 Distinct games for L2
Distinct L2	No Feedback	14 Distinct games for L2 Feedback	No Feedback
Overlapping OPT	7 Overlapping games for OPT 7 Distinct games for OPT	14 Overlapping games for OPT Feedback	7 Overlapping games for OPT 7 Distinct games for OPT
Distinct OPT	No Feedback	14 Distinct games for OPT Feedback	No Feedback
Overlapping BRO	7 Overlapping games for BRO 7 Distinct games for BRO	14 Overlapping games for BRO Feedback	7 Overlapping games for BRO 7 Distinct games for BRO
Distinct BRO	No Feedback	14 Distinct games for BRO Feedback	No Feedback
Overlapping FP	7 Overlapping games for FP 7 Distinct games for FP	14 Overlapping games for FP Feedback	7 Overlapping games for FP 7 Distinct games for FP
Distinct FP	No Feedback	14 Distinct games for FP Feedback	No Feedback

Table 7: Summary of experimental treatments

- The Focal point is a cell with symmetric and high payoffs, having the highest payoff sum, but not necessarily providing the highest payoff to either player (row or column). This design avoids possible confounds with the Optimist strategy. It maximizes social welfare and minimizes social distance.

- Except for the BRO strategy, all non-target heuristic strategies correspond to Row 3. Since it was not possible to place the BRO strategy in Row 3 by construction, we decided to place it systematically in Row 2. This creates a consistent structure across treatments, allowing for comparisons.

- The matrices are designed such that the best reply structure is the same across games.

- The Distinct and Overlapping baseline matrices for each target heuristic are designed to be as similar as possible, sometimes differing by only one or two payoffs.

- The maximin strategy (maximizing the minimum payoff) is another possible non-strategic heuristic that players may adopt. Although this heuristic would not be compatible with the provided information about the algorithm's behavior, we designed games so that the maximin heuristic coincides with the BRO one to avoid confounds. This would only impact our assessment of the distraction effect for the BRO heuristic. In any case, it is worth noting that maximin heuristic players typically constitute a minimal fraction of the population in experimental game settings like ours (Costa-Gomes et al., 2001; Spiliopoulos and Hertwig, 2020).

3.3. Data collection

We conducted our experiment online using Prolific (www.prolific.com), a dedicated platform for experimental tasks. We collected a sample of 827 subjects (average age 38, SD=12, 49% female) from the UK. The sample size for each treatment ranged from 73 to 95 participants. Sample sizes differ among treatments as participants were randomly allocated to them until a minimum number of 70 was reached (the minimum threshold was chosen according to an ex-ante power calculation, see Table D.14; sample sizes are

reported in Table 5). The experiments were incentivized. Participants received a show-up fee of £2 plus a bonus based on their decisions (average bonus £1.6, SD=0.3). Rates were aligned to Prolific standards. The average completion time of the experiment was 16 minutes (SD=7).

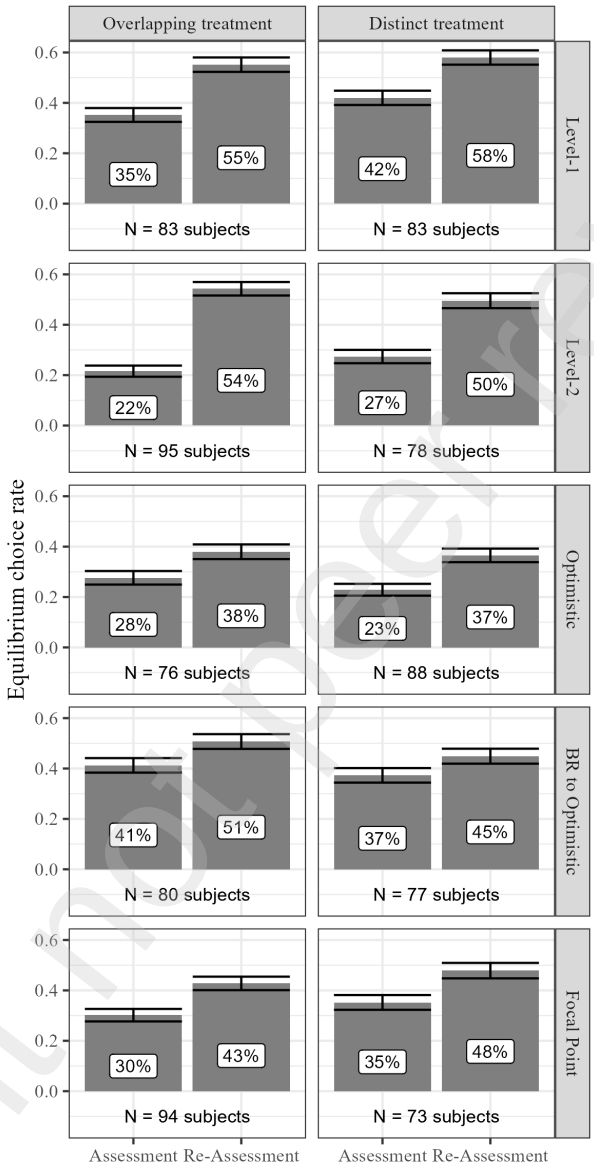
Upon reaching the task page, participants were presented with the informed consent form and instructions. After reading the instructions, participants had to pass a multiple-choice comprehension questionnaire - Appendix B and Appendix C include the instructions and comprehension check. Participants had a maximum of three attempts to answer all questions correctly or were excluded from the experiment. 77 participants failed the comprehension questionnaire. Participants were informed that the experiment included three stages, that they would be playing with a profit-maximizing algorithm assuming they would do the same, and that the algorithm would not adjust its strategy throughout the experiment.

After successfully answering all comprehension questions, participants faced the 14 games of the Assessment stage without receiving feedback. At the end of the Assessment, participants were informed that a new stage would start and that after each game, they would receive feedback about their choice and that of the algorithm, along with the corresponding payoffs. At the end of the Learning stage, participants were informed that the Reassessment stage would begin and that they would not receive any feedback. The games order, as well as the order of the rows and columns of each game, was randomized for each participant. After the experiment, the outcome from three games (one from each stage) was randomly selected to determine the participants' bonuses.

4. Results

4.1. Emergence of learning

As a preliminary question, we wanted to determine whether the Distinct and Overlapping treatments resulted in a significantly increased proportion of equilibrium choices. The equilibrium rate is significantly larger in the Reassessment than in the Assessment (two-sided paired t-test, $p < 0.01$), in the Distinct feedback treatments as well as in the Overlapping feedback treatments - see Table 8 for more details. Results also hold for each of the 10 treatments taken individually, when correcting for multiple testing (correcting with Bonferroni, $p = .05/10 = .005$), except the BRO treatments.



Notes: Error bars give the 95% confidence intervals. Percentages are calculated based on average choice per subject, where each subject makes 14 choices per stage.

Figure 1: Equilibrium choice rates in the Assessment and Reassessment stage

Feedback	Heuristic	Two-sided paired t-test			
		Estimate	Statistic	Degrees of freedom	p-value
All	All	0.160	14.3	826	$< 10^{-3}$
Overlapping	All	0.175	11.5	427	$< 10^{-3}$
Distinct	All	0.144	8.76	398	$< 10^{-3}$
Overlapping	L1	0.200	6.36	82	$< 10^{-3}$
Overlapping	L2	0.327	8.84	94	$< 10^{-3}$
Overlapping	OPT	0.103	3.39	75	0.0011
Overlapping	BRO	0.095	2.71	79	0.0082
Overlapping	FP	0.126	4.44	93	$< 10^{-3}$
Distinct	L1	0.160	5.23	82	$< 10^{-3}$
Distinct	L2	0.222	5.21	77	$< 10^{-3}$
Distinct	OPT	0.136	3.51	87	0.0007
Distinct	BRO	0.076	2.49	76	0.0151
Distinct	FP	0.126	3.23	72	0.0019

Table 8: Difference in equilibrium choice rate between the Assessment and Re-Assessment stages, by treatment

Overall, these initial results show an increased tendency to play the equilibrium strategy, which we will better qualify in the following sections. The average choice rate for each row in each treatment, for the Assessment and Reassessment phase, is reported in Table F.15.

Result 1. *The equilibrium choice rate in the Reassessment is statistically significantly higher than in the Assessment for all treatments (correcting for multiple testing), except only for the BRO target heuristic.*

4.2. Differences in learning

Our data demonstrate that receiving feedback increases the rate of equilibrium choices. However, whether this increase reflects learning a more rational way to play games remains an open question. Looking at Figure 1,

Heuristic treatment	Two-sided two-sample t-test			
	Estimate	Statistic	Degrees of freedom	p-value
All	0.011	0.50	808	0.62
L1	-0.028	-0.64	163	0.52
L2	0.047	0.85	156	0.40
OPT	0.014	0.27	156	0.79
BRO	0.058	1.16	154	0.25
FP	-0.051	-1.06	132	0.29

Table 9: Difference in equilibrium choice rate in the Re-Assessment stage, between the Overlapping and Distinct treatments

there does not appear to be large variations in the equilibrium choice rates comparing the Distinct and Overlapping treatments in the Reassessment. Indeed, a t-test does not reject the null hypothesis that these equilibrium choice rates are equal, be it for each heuristic treatment taken individually or at the aggregate level (see Table 9).

However, it is important to note that in both the Assessment and Re-assessment stages, players encountered a mixture of Overlapping and Distinct types of games. Consequently, players in Overlapping treatments may have improved their performance in the Overlapping-type games (games where they trained during the Learning stage); similarly, players in the Distinct treatments may have enhanced their performance in the Distinct-type games. If this is the case, similar overall rates of equilibrium choices could mask very different types of learning. A first insight is provided by Figure 2, which separately shows the equilibrium choice rates for the Overlapping and Distinct games. The difference in the equilibrium choice rates in the two types of games is evident.

To test this possibility, we compared the rates of equilibrium choices in the Overlapping-type games in the Reassessment stage between the Overlapping and Distinct treatments. The differences are significant, with the equilibrium being chosen significantly more frequently in the Overlapping-type games of the Overlapping treatments (two-sided t-test, Overlapping vs

Game Type	Two-sided two-sample t-test			
	Estimate	Statistic	Degrees of freedom	p-value
Overlapping	0.080	3.04	821	0.002
Distinct	-0.057	-2.20	816	0.028

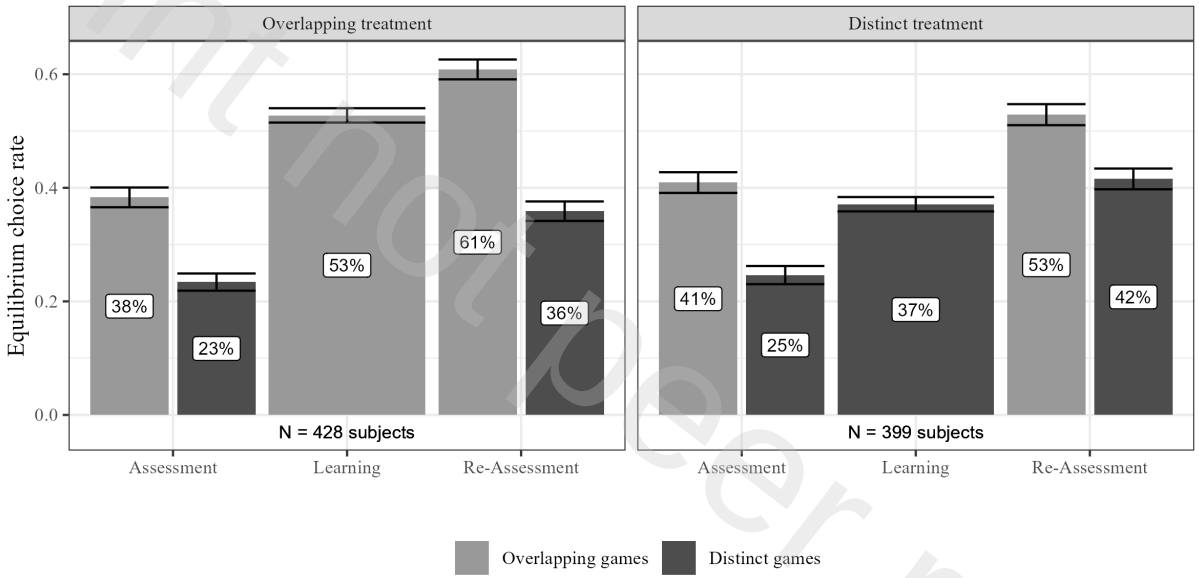
Table 10: Difference in equilibrium choice rate in the Re-Assessment stage, between Overlapping and Distinct treatments

Distinct treatments for Overlapping-type games, $p < 0.01$). The same holds for the Distinct-type games; the equilibrium strategy is chosen significantly more often in the Distinct-type games of the Distinct treatments (two-sided t-test, Overlapping vs Distinct treatments for Distinct-type games, $p = 0.028$, see Table 10).

Result 2. *The frequency of equilibrium choices in the Distinct-type games in the Reassessment is significantly higher in the Distinct treatments than in the Overlapping ones. Vice-versa, the frequency of equilibrium choices in the Overlapping-type games in the Reassessment is significantly higher in the Overlapping treatments than in the Distinct ones.*

Result 2 supports both Hypotheses 1 and 2. To further analyze learning differences across treatments, we conducted four two-way, repeated-measure ANOVA tests. These tests examined the equilibrium choice rate in the Assessment and Reassessment stages, considering the main effects of stage and heuristic (L1, L2, FP, OPT, and BRO), and their interaction. We analyzed the equilibrium rate for four combinations: Overlapping games in Overlapping treatments, Overlapping games in Distinct treatments, Distinct games in Overlapping treatments, and Distinct games in Distinct treatments. This approach helps identify different learning patterns across game types and treatment conditions.

Table 11 presents the test results. Correcting for multiple testing, we consider significant a p = .0125 (Bonferroni correction p = .05/4 = .0125). In the Overlapping treatments (for both Overlapping- and Distinct-type games), we found significant main effects for both heuristic and stage. The heuristic effect demonstrates that different heuristics lead to varying degrees of *distraction from identifying* the equilibrium strategy. The stage effect shows an



Notes: The outcome variable is the equilibrium choice rate at the (subject, stage) level. The number of observations per vertical bar equals the number of participants in the corresponding feedback treatment groups.

Figure 2: Equilibrium choice rates in the Overlapping and Distinct treatment groups

	Dependent variable: Equilibrium choice rate			
	(1)	(2)	(3)	(4)
Feedback Treatment	Overlapping	Overlapping	Distinct	Distinct
Game Type	Overlapping	Distinct	Overlapping	Distinct
Stage	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p < 0.001$
Heuristic	$p < 0.001$	$p = 0.001$	$p < 0.001$	$p = 0.047$
Heuristic × Stage	$p < 0.001$	$p < 0.001$	$p = 0.017$	$p = 0.489$
Observations	856	856	764	764
Subjects	428	428	382	382

Notes: The Heuristic variable has five levels (L1, L2, OPT, BRO, and FP). The Stage variable has two levels (Assessment and Reassessment). The outcome variable is the equilibrium choice rate at the (subject, stage, game type) level. The number of observations per column equals the number of stages × the number of participants in the treatment group.

Table 11: Two-way ANOVA. Difference of equilibrium choice by stage across feedback and heuristic treatments

overall increase in equilibrium choice rates from Assessment to Reassessment, indicating *learning*. Additionally, the significant interaction effect between heuristic and stage suggests that different heuristics lead to varying degrees of *distraction from learning* the equilibrium strategy.

For the Distinct treatments, we observed different patterns. In Overlapping-type games, the rate of equilibrium choices increased significantly across stages (significant stage effect). The heuristic main effect was significant, indicating that players' ability to play equilibrium varies across target heuristics. The interaction effect between heuristic and stage albeit low, was not significant if correcting for multiple testing, showing a tension between the desire of players to follow the heuristic and learning the equilibrium.

Regarding Distinct-type games in the Distinct treatment, only the stage main effect was significant, highlighting an increase in equilibrium choice rates across stages. The lack of significant effects for the heuristic main effect and the stage::heuristic interaction is expected given our experimental design. In these settings, the Distinct-type games observed in all three stages (Assessment, Learning, and Reassessment) are structurally identical across

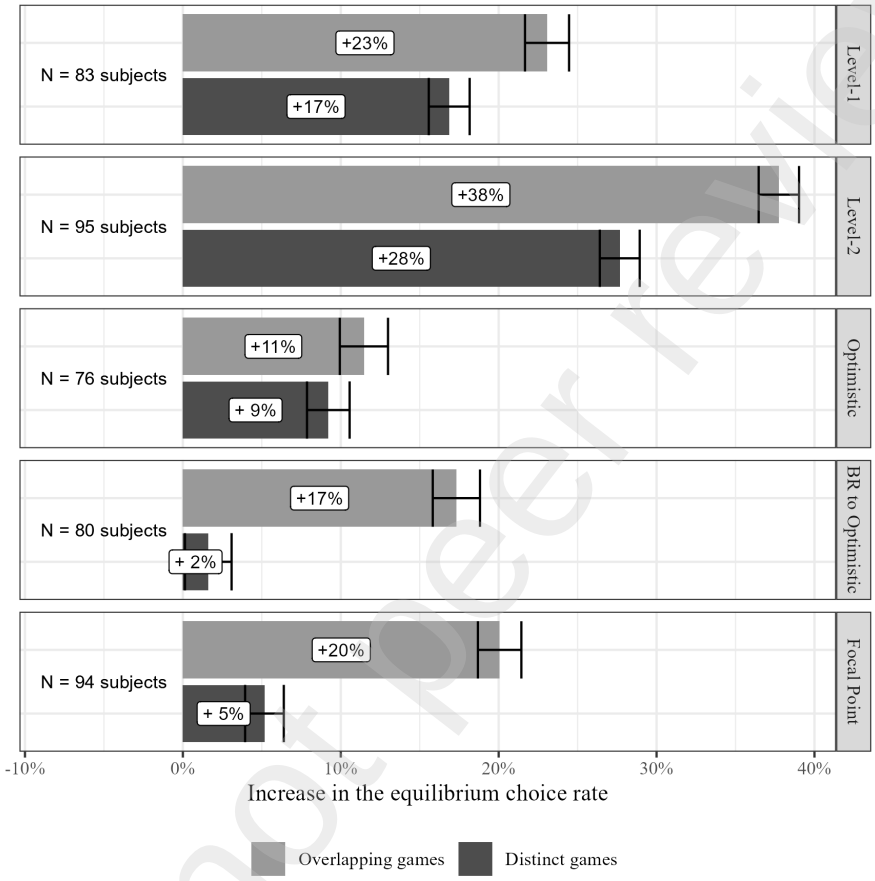
treatments, and feedback information consistently reinforces the equilibrium choice strategy. For this reason, we did not expect to observe any effect due to the nature of the heuristics.

Overall, the results from the ANOVA strongly support both Hypotheses 1 and 2.

4.3. Heterogeneity in learning outcomes

Results so far suggest that the presence of the target heuristics greatly affects the learning process. It remains to be discussed how the different heuristics interfere with learning. To address this point, we focus only on the Overlapping treatments, and observe how learning rates differ by target heuristic and game type. To measure learning, we compute the difference between the rate of equilibrium choices in the Reassessment and Assessment stages separately for Overlapping and Distinct games. These data allow us to verify a) how learning differed across game types, and b) the extent to which each heuristic was hindering the learning of the equilibrium strategy.

Figure 3 reports the increase of equilibrium rate for the five different heuristics, in the Distinct- and Overlapping-type games separately. We see large differences across treatments in two dimensions: first, how much the equilibrium rate increases; second, how large is the difference between the increase in equilibrium choices in Overlapping and Distinct games. The increase in equilibrium rate varies from a minimum of 2% (BRO Distinct), to a maximum of 38% (L2 Overlapping); while the difference spans from 2% (OPT) to 15% (FP and BRO). In all cases the rate of equilibrium choices increases more in the Overlapping games, suggesting that indeed some subjects learn to apply the target heuristic rather than equilibrium play. Nonetheless, the impact of the heuristics is very diverse. The BRO and FP treatments show similar effects. In both, we observe a large increase of choices compatible with both the equilibrium and the heuristic in the Overlapping treatments (17% for BRO and 20% for FP), but a very small increase of equilibrium choices in Distinct treatments (2% for BRO and 5% for FP). This suggests that both heuristics are widely recognized and adopted by players, and tend to distract from the selection of the equilibrium choice when reinforced. Conversely, the OPT heuristic is not much recognized and adopted by players, as the increase in equilibrium rate is similar across the two types of games. Both L1 and L2 strategies are observable by some players (an increase of 6% in equilibrium choices in Overlapping games for L1, and an increase of 10% for L2), but do not prevent other players from learning the equilibrium



Notes: The sample is restricted to subjects in the Overlapping treatment. Each bar illustrates the change in the rate of equilibrium choice from the Assessment to the Reassessment stage. The bar color differentiates between Overlapping and Distinct game types. For instance, in the Overlapping L1 treatment, equilibrium choice rates in Overlapping-type games increased from 18% in the Assessment to 35% in the Reassessment, resulting in a 17% difference. The number of observations per horizontal bar equals the number of participants in the corresponding feedback and treatment groups.

Figure 3: Increase of the equilibrium choice rates by game type & heuristic, in Overlapping treatments

Heuristic treatment	Two-sided paired t-test			
	Estimate	Statistic	Degrees of freedom	p-value
All	0.100	5.28	427	$< 10^{-3}$
L1	0.062	1.53	82	0.129
L2	0.101	2.85	94	0.005
OPT	0.023	0.52	75	0.605
BRO	0.157	3.34	79	0.001
FP	0.149	3.27	93	0.002

Table 12: Difference in equilibrium rate between Overlapping and Distinct games for Overlapping treatments

strategy (17% for L1 and 28% for L2). Table 12 compares systematically the difference in equilibrium rate between the Overlapping and Distinct feedback treatment for each given heuristic using a two-sided two-sample t-test. This leads to Result 3, which supports Hypothesis 3:

Result 3. *When heuristic and equilibrium strategies overlap, the strength with which the heuristic interferes with the learning of equilibrium varies depending on the type of heuristic.*

4.4. What do subjects learn?

These results suggest that feedback from Overlapping-type games encourages some participants to learn simpler heuristics rather than the more cognitively demanding equilibrium strategy. However, the real impact of heuristics remains to be investigated. How often subjects prefer to solve a complex situation through the use of heuristics, or whether people that make use of heuristics are willing to give them up remains to be seen. To generalize our claims, we designed a more precise measure of heuristic play, which we call the "target-heuristic index". The index is so calculated: For each player, we approximated the tendency to select the target heuristic by summing the rates of R1 choices in Overlapping-type games and R2 choices in Distinct games (the rows where the target heuristics lay). We then debiased this sum by subtracting: a) The rate of R1 choices in Distinct-type

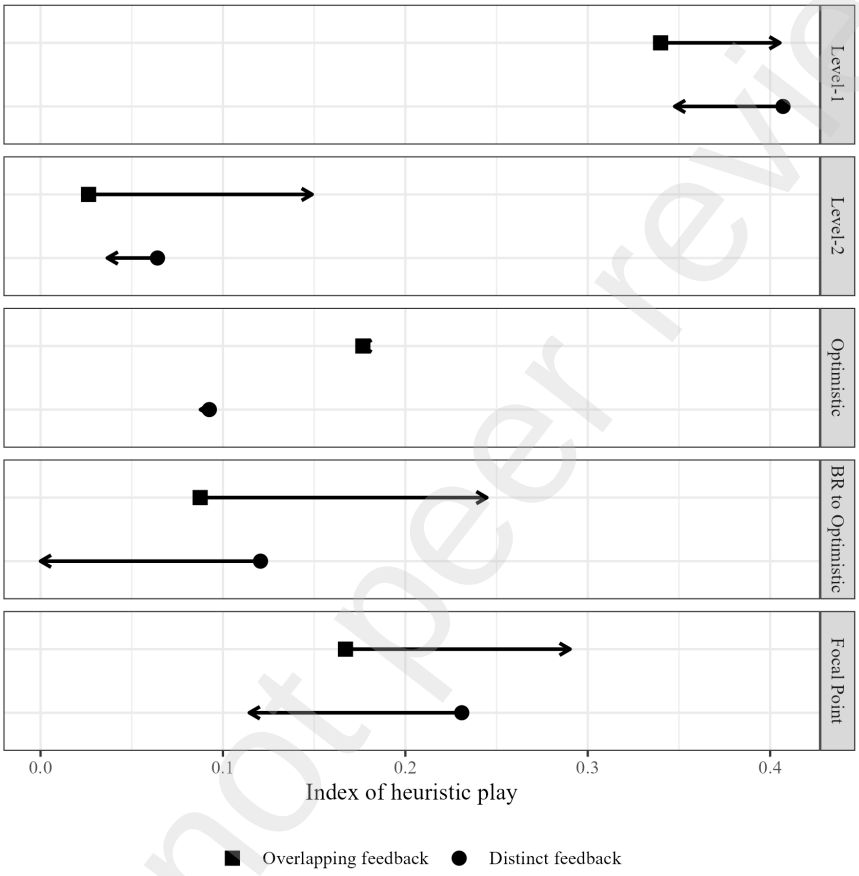
games, assuming players who selected R1 in Distinct-type games behaved similarly in Overlapping-type ones, aiming, e.g., for the Equilibrium. b) The frequency of R2 choices in Overlapping games, assuming agents who chose R2 in Overlapping games may have done so consistently in Distinct-type games as well[5]. The index is then divided by 2 for normalization. Thus, our measure of compatibility of choices with a target heuristic is defined as follows:

$$\begin{aligned}\text{Target-heuristic index} = (\ & f_{\text{R1 in Overlapping games}} + f_{\text{R2 in Distinct games}} \\ & - f_{\text{R1 in Distinct games}} - f_{\text{R2 in Overlapping games}}\ ) \ / \ 2\end{aligned}$$

Figure 4 offers a dynamic picture of how heuristics are played before and after learning has occurred. It illustrates how the tendency to play a target heuristic is affected by training on Overlapping and Distinct types of games. The arrow origins (square for treatments with training with Overlapping games, circle with Distinct ones) show the rate of heuristic-based choices in the Assessment stage. The arrow tips indicate the rate of heuristic-based choice after the learning stage, with the direction showing whether the use of heuristics increased or decreased.

The ex-ante rates of heuristic-based choices vary significantly across different heuristics. Unsurprisingly, FP and L1 are the most commonly used heuristics by players initially (approximately 20% and 37% on average, respectively). These strategies are readily detectable and do not require complex considerations about the opponent's behavior. Instead, the BRO and OPT heuristics are used less frequently (respectively 10% and 12%, on average). This is reasonable, as both are based on the assumption that at least one player will select a strategy because it includes the largest payoff. This rather naive approach is likely adopted only by players with a high degree of myopia and egocentrism, making these heuristics rarely observed in previous studies (Spiliopoulos and Hertwig, 2020). Indeed, while players may not fully understand the optimal course of action, they generally recognize that blindly pursuing a payoff advantageous only to themselves is not a reasonable choice (and do not expect such behavior from their opponent). The reasoning behind the FP choice is more sophisticated: the payoffs are advantageous to both players, making it possible to expect that the counterpart will make

[5]Such a choice can be driven by applying the BRO or Maximin heuristic, which always lie in Row 2, or any other rationale we cannot envisage.



Notes: The starting point of the arrow (either a square or a circle) gives the index value in the Assessment stage. The tip of the arrow gives the index value in the Re-Assessment stage. The shape of the starting point of the arrow indicates the feedback treatment group of the participant. Ex: in the Overlapping L1 treatment, the index value was around 0.34 in the Assessment stage and around 0.41 in the Re-Assessment stage. The number of observations per treatment and stage equals the number of participants in the corresponding group.

Figure 4: Learning-induced change in the index of heuristic play

the same choice. L2 is the least commonly used strategy ex-ante (less than 5%). This heuristic requires sophisticated strategic reasoning (implying to best respond to an L1 opponent) (Nagel, 1995; Stahl and Wilson, 1995; Gill and Prowse, 2016), and is typically adopted less frequently, especially in interactive contexts like ours, before learning takes place.

Figure 4 shows how, in all treatments, the rate of heuristic-based choices systematically increased in the Overlapping treatment group and decreased in the Distinct ones, except the OPT treatment, which remained substantially unchanged. This implies that in Overlapping treatments, players tended to switch to using the heuristic after receiving feedback, whereas in the Distinct treatment group, several players switched from the heuristic to the equilibrium strategy. The largest changes in strategy are observed in the FP and BRO treatments, where both heuristics are easily detected when feedback supports them, but also clearly sub-optimal when not coinciding with equilibrium play. Furthermore, L2 is a strategy that increases significantly when supported by feedback, showing that people improve their level of strategic thinking when trained (aligned with results of Marchiori et al. (2021)). OPT and L1 instead are rather stable, for very different reasons. As mentioned above, OPT is a rarely observed heuristic, while L1 is by construction a safe choice under any circumstance, when there is no certainty about the behavior of the counterpart (Spiliopoulos and Hertwig, 2020).

To test our observations, we grouped subjects by the type of training they received (Overlapping vs. Distinct), independently of the target heuristic, and compared their index of heuristic play between the Assessment and Reassessment stages (see Table 13). The index of heuristic play is significantly larger after receiving Overlapping feedback (two-sided paired t-test, $p < 0.001$) and significantly lower after receiving Distinct feedback (two-sided paired t-test, $p < 0.001$). We also tested whether this general pattern holds for each heuristic taken individually. As suggested by Figure 4, the index of heuristic play is increased (decreased) after receiving Overlapping (Distinct) feedback, in all treatments but the OPT. These variations are significant for the BRO ($p < 0.001$ for Overlapping, and $p = 0.018$ for Distinct), FP ($p = 0.002$ for Overlapping, and $p = 0.009$ for Distinct), and Overlapping L2 ($p < 0.001$) treatments. We can conclude that:

Result 4. *Relative to the Assessment stage, the frequency of heuristic-consistent choices in the Reassessment is higher after learning on Overlapping-type games and lower after learning on Distinct-type ones.*

Feedback	Heuristic	Two-sided paired t-test			
		Estimate	Statistic	Degrees of freedom	p-value
All	All	0.019	1.47	826	0.1412
Overlapping	All	0.096	5.47	427	$< 10^{-3}$
Distinct	All	-0.063	-3.43	398	$< 10^{-3}$
Overlapping	L1	0.065	1.61	82	0.1104
Overlapping	L2	0.122	3.56	94	$< 10^{-3}$
Overlapping	OPT	-0.001	-0.02	75	0.9806
Overlapping	BRO	0.157	3.71	79	$< 10^{-3}$
Overlapping	FP	0.123	3.12	93	0.0024
Distinct	L1	-0.059	-1.24	82	0.2199
Distinct	L2	-0.027	-0.91	77	0.3660
Distinct	OPT	-0.005	0.15	87	0.8785
Distinct	BRO	-0.121	-2.42	76	0.0179
Distinct	FP	-0.116	-2.69	72	0.0090

Table 13: Difference in index of heuristic choice between the Assessement and Re-Assessment stages

Considering Results 1, 2, and 4, we confirm both Hypotheses 1 and 2.

5. Discussion and conclusion

Our study challenges the assumption that providing complete feedback in repeated strategic interactions necessarily leads to an increase in rational play. We argue that even when feedback is provided and there is no uncertainty about the game structure and incentives, the structure of the game itself can introduce ambiguity that influences players' decision strategies.

Our experimental design introduces a novel approach by constructing a set of games with theoretically similar structures. We categorize these games into two types: Overlapping, where the equilibrium action aligns with a heuristic strategy, and Distinct, where it does not. We hypothesized that feedback in Overlapping games would reinforce heuristic play rather than equilibrium strategies, despite providing accurate and unambiguous information. Notably, from a theoretical standpoint, in all our games, feedback information consistently reinforces the equilibrium action. However, our hypothesis suggests that players' interpretation and use of this feedback may vary depending on the game structure.

Our findings support this hypothesis. Consistently with Hypothesis 1, when learning with Overlapping-type games, players' behavior aligns more closely with heuristic play than with equilibrium strategies (Results 2, 3, and 4). We attribute this to the combination of near-optimal returns and lower cognitive demands associated with heuristic play, factors that are widely considered to contribute to the ecological validity of these heuristics (Hertwig and Ortmann, 2001; Todd and Gigerenzer, 2012; Mousavi and Kheirandish, 2014; Callaway et al., 2021).

Conversely, confirming Hypothesis 2, feedback from Distinct-type games leads to behavioral outcomes more consistent with rational play (Results 2 and 3). However, the highest rate of equilibrium-compatible choices is not always induced by training with distinct games. Under certain (heuristic-dependent) circumstances, it is possible that training with overlapping games induces a higher rate of equilibrium-compatible choices. Thus, the effectiveness of feedback in promoting equilibrium play ultimately depends on the joint effect of equilibrium and heuristic strategy.

Interestingly, we observed variations in the "distraction from equilibrium" effect across different heuristics (Hypothesis 3 supported by Results 3 and 4). Whereas this effect is generally present and statistically significant for

most heuristics, its magnitude varies, suggesting an interplay between game structure and specific heuristic strategies.

Our results have two key implications for future behavioral game theory research and the interpretation of past studies. First, we extensively show that an increase in equilibrium actions does not necessarily indicate an increased level of rationality or an improved ability to analyze a game. Players may simply have learned a heuristic that happens to align with the equilibrium strategy. Second, the game structure itself can be a source of ambiguity in repeated strategic interactions, significantly influencing what players learn.

These insights suggest that to effectively induce learning of rational play, it may be more beneficial to design situations where the equilibrium choice is incompatible with common heuristics. This approach would be particularly valuable, for example, when policymakers aim to encourage agents to generalize rational strategies to similar situations. Such generalization could enhance decision-making across various contexts. Conversely, if the specific learning outcome is less critical and the focus of a policy is instead that of inducing a desired behavior in a specific strategic setting, such an outcome could be better and more easily achieved by designing the strategic setting so that the desired choice behavior overlaps with some salient heuristics.

In conclusion, our study highlights the relationship between feedback, game structure, and learning outcomes in strategic interactions. We strongly underscore the need for careful consideration of these factors in both research design and the interpretation of results from past studies.

Acknowledgments

To be filled before publication.

Funding sources

Funded by: the European Union – Next Generation EU, project titled "Strategic thinking development in an ever changing world", P2022TALJF - PRIN 2022 PNRR M4C2 I. 1.1; the Independent Research Fund Denmark, project titled "A behavioral investigation of the determinants of strategic learning", project number: 1028-00051B; the Independent Research Fund Denmark, project titled "Unraveling transfer of knowledge in economic games", project number: 0166-00005A. The funding sources were not directly involved in any of the stages of the research.

Declaration of generative AI and AI-assisted technologies in the writing process

The authors did not use generative AI during the process of scientific writing.

References

Callaway, F., Griffiths, T.L., Rehbinder, G.K., 2021. Rational Heuristics for One-Shot Games. Working Paper .

Camerer, C.F., 2011. Behavioral Game Theory: Experiments in Strategic Interaction. Princeton University Press.

Camerer, C.F., Ho, T.H., Chong, J.K., 2004. A Cognitive Hierarchy Model of Games. The Quarterly Journal of Economics 119, 861–898. doi:10.1162/0033553041502225.

Cooper, R.W., DeJong, D.V., Forsythe, R., Ross, T.W., 1990. Selection Criteria in Coordination Games: Some Experimental Results. The American Economic Review 80, 218–233. arXiv:2006744.

Costa-Gomes, M., Crawford, V.P., Broseta, B., 2001. Cognition and Behavior in Normal-Form Games: An Experimental Study. Econometrica 69, 1193–1235. doi:10.1111/1468-0262.00239.

Costa-Gomes, M.A., Weizsäcker, G., 2008. Stated Beliefs and Play in Normal-Form Games. The Review of Economic Studies 75, 729–762. doi:10.1111/j.1467-937X.2008.00498.x.

Crawford, V.P., Gneezy, U., Rottenstreich, Y., 2008. The Power of Focal Points Is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures. American Economic Review 98, 1443–1458. doi:10.1257/aer.98.4.1443.

Devetag, G., Di Guida, S., Polonio, L., 2016. An eye-tracking study of feature-based choice in one-shot games. Experimental Economics 19, 177–201. doi:10.1007/s10683-015-9432-5.

Erev, I., Haruvy, E., 2016. 10. Learning and the Economics of Small Decisions, in: 10. Learning and the Economics of Small Decisions. Princeton University Press, pp. 638–716. doi:10.1515/9781400883172-011.

Erev, I., Roth, A.E., 1998. Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria. The American Economic Review 88, 848–881. arXiv:117009.

Fehr, E., Schmidt, K.M., 1999. A Theory of Fairness, Competition, and Cooperation. The Quarterly Journal of Economics 114, 817–868. doi:10.1162/003355399556151.

Fudenberg, D., Levine, D.K., 1998. The Theory of Learning in Games. MIT Press.

Fudenberg, D., Levine, D.K., 2009. Learning and Equilibrium. Annual Review of Economics 1, 385–420. doi:10.1146/annurev.economics.050708.142930.

Gill, D., Prowse, V., 2016. Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A Level-k Analysis. Journal of Political Economy 124, 1619–1676. doi:10.1086/688849.

Goeree, J.K., Holt, C.A., 2001. Ten Little Treasures of Game Theory and Ten Intuitive Contradictions. American Economic Review 91, 1402–1422. doi:10.1257/aer.91.5.1402.

Hertwig, R., Ortmann, A., 2001. Experimental practices in economics: A methodological challenge for psychologists? Behavioral and Brain Sciences 24, 383–403. doi:10.1017/S0140525X01004149.

Ho, T.H., Camerer, C.F., Chong, J.K., 2007. Self-tuning experience weighted attraction learning in games. Journal of Economic Theory 133, 177–198. doi:10.1016/j.jet.2005.12.008.

Jackson, M.O., Xing, Y., 2014. Culture-dependent strategies in coordination games. Proceedings of the National Academy of Sciences 111, 10889–10896. doi:10.1073/pnas.1400826111.

Marchiori, D., Di Guida, S., Polonio, L., 2021. Plasticity of strategic sophistication in interactive decision-making. Journal of Economic Theory 196, 105291. doi:10.1016/j.jet.2021.105291.

Marchiori, D., Warglien, M., 2008. Predicting Human Interactive Learning by Regret-Driven Neural Networks. Science 319, 1111–1113. doi:10.1126/science.1151185.

Mehta, J., Starmer, C., Sugden, R., 1994. Focal points in pure coordination games: An experimental investigation. Theory and Decision 36, 163–185. doi:10.1007/BF01079211.

Mousavi, S., Kheirandish, R., 2014. Behind and beyond a shared definition of ecological rationality: A functional view of heuristics. Journal of Business Research 67, 1780–1785. doi:10.1016/j.jbusres.2014.03.004.

Nagel, R., 1995. Unraveling in Guessing Games: An Experimental Study. The American Economic Review 85, 1313–1326. arXiv:2950991.

Parravano, M., Poulsen, O., 2015. Stake size and the power of focal points in coordination games: Experimental evidence. Games and Economic Behavior 94, 191–199. doi:10.1016/j.geb.2015.05.001.

Polonio, L., Coricelli, G., 2019. Testing the level of consistency between choices and beliefs in games using eye-tracking. Games and Economic Behavior 113, 566–586. doi:10.1016/j.geb.2018.11.003.

Polonio, L., Di Guida, S., Coricelli, G., 2015. Strategic sophistication and attention in games: An eye-tracking study. Games and Economic Behavior 94, 80–96. doi:10.1016/j.geb.2015.09.003.

Salmon, T.C., 2001. An Evaluation of Econometric Models of Adaptive Learning. Econometrica 69, 1597–1628. doi:10.1111/1468-0262.00258.

Schelling, T.C., 1960. The Strategy of Conflict. First edition ed., Harvard University Press, Cambridge, Mass.

Selten, R., Abbink, K., Buchta, J., Sadrieh, A., 2003. How to play (3×3)-games.: A strategy method experiment. Games and Economic Behavior 45, 19–37. doi:10.1016/S0899-8256(02)00528-6.

Selten, R., Chmura, T., 2008. Stationary Concepts for Experimental 2x2-Games. American Economic Review 98, 938–966. doi:10.1257/aer.98.3.938.

Spiliopoulos, L., Hertwig, R., 2020. A map of ecologically rational heuristics for uncertain strategic worlds. Psychological Review 127, 245–280. doi:10.1037/rev0000171.

Stahl, D.O., Wilson, P.W., 1995. On Players' Models of Other Players: Theory and Experimental Evidence. Games and Economic Behavior 10, 218–254. doi:10.1006/game.1995.1031.

Todd, P.M., Gigerenzer, G., 2012. Ecological Rationality: Intelligence in the World. Oxford University Press, USA.

Van Lange, P.A.M., 1999. The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. Journal of Personality and Social Psychology 77, 337–349. doi:10.1037/0022-3514.77.2.337.

Van Lange, P.A.M., 2000. Beyond Self-interest: A Set of Propositions Relevant to Interpersonal Orientations. European Review of Social Psychology 11, 297–331. doi:10.1080/14792772043000068.

Weber, R.A., 2003. 'Learning' with no feedback in a competitive guessing game. Games and Economic Behavior 44, 134–144. doi:10.1016/S0899-8256(03)00002-2.

Appendix A. Ethical approval

The experiment was approved by the "Research Ethics Committee" of the University of Southern Denmark, Case nr. 20/39561.

Appendix B. Instructions

Please read these instructions carefully and then press the button on the bottom of this page to proceed

This study includes three parts. In each part, you will face a sequence of interactive decisions in which you will be asked to choose among three possible actions. "Interactive" means that the outcome of your decisions depends on both your choice and that of a counterpart.

In all decisions, your counterpart will be the "Computer", that is, an algorithm that uses a pre-determined decision rule. You and the Computer will choose simultaneously, meaning that at each trial you and the Computer will take your own decision without knowing in advance that of the other.

The Computer will choose as to gain as many points as possible under the assumption that you will do the same, and will adopt the same decision rule throughout the whole experiment. Thus, the Computer will not adjust its strategy of action to your previous choices.

Each interactive decision, which we will refer to as a game, will be represented in a table, as shown below.

Game

		Computer's actions		
		Action 1	Action 2	Action 3
Your actions	Action 1	31 27	38 36	40 47
	Action 2	43 34	26 31	31 32
	Action 3	33 55	46 29	35 37

In the games you will play in this study, you and the Computer can select among three possible actions labeled "Action 1", "Action 2", and "Action 3" (see the game above for an illustration). Your actions are displayed by row, and those of the Computer by column. To each combination of actions by you and the Computer, there corresponds a cell of the table that includes a pair of numbers: the first number (in red) is your payoff in experimental points, and the second (in blue) the payoff of the Computer.

To facilitate understanding, hovering your mouse over your and the Computer's actions highlights cells in the corresponding row and column, and hovering over a cell highlights the combination of actions that yields it. You can select one of your actions by clicking on one of the action labels. You will not be allowed to get back to previous games and revise your actions.

Example: Referring to the game illustrated above, if you choose Action 1 and the Computer chooses Action 3, you will get 47 points (in red) and the

Computer 40 points (in blue). You can check this by hovering the mouse pointer over the game cell with payoffs 47 and 40. Earnings

Your performance in this experiment will be assessed based on the points you get with your choices.

Beside your participation fee of £2, you will get a bonus based on the outcome of your decisions in three randomly sampled games (one for each part). One point will be converted to £0.01.

Example: Suppose that in the three games randomly selected, your payoff was 31 points, 62 points, and 44 points. Therefore, you will be paid a bonus of £0.31 + £0.62 + £0.44 = £1.37. Important points to remember

- *With your decisions you choose a row of the game, whereas the Computer chooses one of the columns.*
- *Only the combination of the row and column choices determines your payoff and that of the Computer.*
- *Neither you nor the Computer will know in advance the choice of the counterpart.*
- *The Computer will choose as to maximize its own payoff, under the assumption that you will do the same.*
- *The Computer will use the same decision rule throughout the whole experiment.*
- *You will receive feedback about the computer actions only in Part 2 of the study. In Part 1 and 3 you will make your choices without being informed about what the Computer has chosen.*
- *The more points you accumulate with your decisions, the larger your final bonus.*

Appendix C. Comprehension check

Please answer the following questions and press the button on the bottom of this page to proceed

Game

		Computer's actions Action 1	Computer's actions Action 2	Computer's actions Action 3
Your actions	Action 1	31 / 27	38 / 36	40 / 47
	Action 2	43 / 34	26 / 31	31 / 32
	Action 3	33 / 55	46 / 29	35 / 37

1) In the game on the left, if you choose Action 3:

- ☐ *You get 55 points*
- ☐ *You get 29 points if the Computer chooses Action 2*
- ☐ *The Computer gets its largest payoff*
- ☐ *Don't know*

2) In the game on the left, if you choose Action 1 and the Computer chooses Action 3, then:

- ☐ *You get 55 points and the Computer 33*
- ☐ *You get 37 points and the Computer 35*
- ☐ *You get 47 points and the Computer 40*
- ☐ *Don't know*

3) In the game on the left, if you choose Action 2 and the Computer chooses Action 1, then:

- ☐ *You get 34 points*
- ☐ *The Computer gets 34 points*
- ☐ *You get 36 points*
- ☐ *Don't know*

Appendix D. Power analysis

Input	Tail(s)	One-sided
	Effect size d	0.5
	α error probability	0.05
	Power ($1 - \beta$ error probability)	0.9
	Allocation ratio N2/N1	1
Output	Noncentrality parameter δ	2.958
	Critical t	1.656
	Degrees of freedom	138
	Sample size group 1	70
	Sample size group 2	70
	Total sample size	140
	Actual power	0.903

Table D.14: A priori power analysis for choosing the sample size

Appendix E. Data processing

In total, 77 subjects failed to answer correctly all questions of the comprehension check three times and were not allowed to continue with the experimental task.

Of all the data collected, the first subjects participating in the FP treatments were excluded. A short time after we started the data collection, we realized that, in the FP treatment, the BRO heuristic was not lying in row 2, as it should have. For this reason, we modified the game and used the correct version for all the following sessions, but had to remove the first data collected for the FP treatments.

Appendix F. Additional tables

Treatment	Stage 1 Assessment			Stage 3 Reassessment			Sample size
	R1	R2	R3	R1	R2	R3	
Overlapping L1	0.35	0.40	0.25	0.55	0.32	0.13	N = 83 subjects
Distinct L1	0.42	0.39	0.19	0.58	0.28	0.14	N = 83 subjects
Overlapping L2	0.22	0.45	0.34	0.54	0.26	0.20	N = 95 subjects
Distinct L2	0.27	0.36	0.37	0.50	0.27	0.23	N = 78 subjects
Overlapping OPT	0.28	0.33	0.40	0.38	0.23	0.39	N = 76 subjects
Distinct OPT	0.23	0.40	0.37	0.36	0.24	0.40	N = 88 subjects
Overlapping BRO	0.41	0.25	0.34	0.51	0.23	0.26	N = 80 subjects
Distinct BRO	0.37	0.22	0.41	0.45	0.13	0.42	N = 77 subjects
Overlapping FP	0.30	0.32	0.38	0.43	0.22	0.35	N = 94 subjects
Distinct FP	0.35	0.30	0.35	0.48	0.15	0.37	N = 73 subjects

Notes: Choice rates for the row player's actions Row 1 (R1), Row 2 (R2), and Row 3 (R3) by stage and treatment group. Because of rounding, they may sum up to 1.01.

Table F.15: Row player's choice rates by stage and treatment

Appendix G. Games

5.3 6.4	6 **5.4**	6.5 6.6
5.5 6.3	5.1 7.4	**7** **5.2**
8 8	9 5	**7.5** 6.2

5.3 6.4	6 **5.2**	6.5 6.6
5.5 6.3	5.1 7.4	**6.1** **5.4**
8 8	9 5	**7** 6.2

Table G.16: Example of Distinct- and Overlapping-type payoff matrices for the L2 heuristic

6.4 **5.5**	5.3 6	6.9 7
6.6 6.1	5 6.8	**8** **5.4**
7.5 7.5	**7.6** 5.2	5.1 6.3

5.5 **5.1**	5 6	6.1 7.1
6.6 **6.5**	4.5 7.2	7 6.3
8 8	9 4	5.3 5.6

Table G.17: Example of Distinct- and Overlapping-type payoff matrices for the "Best-Response to Optimistic" heuristic

6.3 **5.3**	6.5 6.2	6.8 **7.7**
6.7 6.7	5 **7.6**	6.6 **5.2**
7.5 7.5	7.6 5.1	6.1 7.2

6.3 **5.2**	6.5 6.2	6.8 **7.6**
6.7 6.7	5 **7.7**	6.6 **5.3**
7.5 7.5	7.6 5.1	6.1 7.2

Table G.18: Example of Distinct- and Overlapping-type payoff matrices for the "Optimistic" heuristic

5.5 **6.5**	5 6	6.1 7.1
6.6 **5.1**	4.5 7.2	7 6.3
8 8	9 4	5.3 5.6

5.5 **5.1**	5 6	6.1 7.1
6.6 **6.5**	4.5 7.2	7 6.3
8 8	9 4	5.3 5.6

Table G.19: Example of Distinct- and Overlapping-type payoff matrices for the "Focal Point" heuristic

	L1		L2		OPT		BRO		FP	
	O	D	O	D	O	D	O	D	O	D
L1	T	M	B	B	B	B	B	B	B	B
L2	B	B	T	M	B	B	B	B	B	B
OPT	B	B	B	B	T	M	B	B	B	B
BRO	M	M	M	M	M	M	T	M	M	M
FP	B	B	B	B	B	B	B	B	T	M
EQ	T	T	T	T	T	T	T	T	T	T

Notes: Letters T, M and B stand for "Top", "Middle" and "Bottom". Letters O and D refers to "Overlapping" and "Distinct" matrices. These notations specify a matrix and row in Tables G.16 to G.19. Rows indicate which strategies are consistent with the action in the case.

Table G.20: Relation between choices and strategies for each payoff matrix