Game-form recognition in dynamic interactions*

Konrad Grabiszewski† Alex Horenstein ‡

Abstract

Succeeding at game-form recognition is critical for strategy selection and, ultimately success or failure when playing games. Whether people understand the games they participate in is the empirical question addressed in this article. Replicated in 27 different strategic settings with 4,788 subjects, our test shows that between 12% and 65% of subjects fail at game-form recognition. Additionally, in our setup, we find that game-form recognition is more challenging than backward induction. To collect data, we conducted a mobile experiment using *Blues and Reds*, an app available globally for free on iOS and Android devices.

JEL Codes: C72, C99, D91.

Key words: game-form recognition, finite dynamic games, mobile experiment.

*This article was previously circulated under the title "A mobile experiment on tree construction." We are grateful to Andrew Leone, who as the Vice Dean for Faculty Development and Research at the University of Miami Business School, helped us secure financing to conduct this project. For their comments, suggestions, and feedback, we thank Raphael Boleslavsky, Hector Chade, Christopher Cotton, Kfir Eliaz, Simon Grant, John Hey, Alejandro Manelli, Ariel Rubinstein, Edward Schlee, and Emanuel Vespa.

†Mohammad bin Salman College, Saudi Arabia; konrad.grabiszewski@gmail.com

‡Department of Economics, University of Miami, FL, USA; horenstein@bus.miami.edu

1 Introduction

Game-form recognition success is a crucial skill that enables players to select optimal strategies as they must understand the game they participate in (Chou et al. (2009), Cason and Plott (2014)). The problem of game-form recognition is of paramount concern to the foundations of game theory. From a theoretical perspective, equilibrium analysis requires that a model describing an interaction is known to the players participating in the interaction (Myerson (1991)). From an experimental perspective, testing game theory requires that subjects correctly understand the interaction being studied (Chou et al. (2009)).

The task of creating a mental model of an interaction is neither simple nor natural. Consequently, whether we can assume that people succeed at game-form recognition is ultimately an empirical question and the main question that this paper addresses. Our focus is on the players' understanding of finite dynamic games with perfect and complete information. That is, games representable as trees and solvable by backward induction.

A player correctly recognizes a dynamic game if they understand who the players are, what their strategies and payoffs are, and what the sequence of choices is. Of particular interest is the problem of players comprehending the links between their choices and their payoffs (Cason and Plott (2014), Bull et al. (2019), Drichoutis and Nayga (2020)).

Testing whether people correctly understand the game they play is not trivial as their perception remains in their minds and is not directly observable. A typical dataset – as well as the one used in this article – includes the subject's choices but not their mental process. To solve the non-observability problem, our test of game-form recognition consists of two stages and resembles the experimental design from Levitt et al. (2011).

In the first stage, the subject plays a "tree interaction," which resembles a standard game-theoretic tree. The subject either wins or loses. Winning requires correct backward induction. The sole purpose of this stage is to select the backward-inducting (BI) subjects; that is, subjects with the ability to backward induct.

In the second stage, the subject plays a "non-tree interaction." Again, there are only two possible outcomes: the subject either wins or loses. However, now, winning is more challenging as it requires not only correct backward induction but also correct game-form recognition.

Importantly, the tree and non-tree interactions constitute a pair of analogous interactions: they share the same graph and game-theoretic properties. Consequently, what separates the BI subjects selected in a given tree from winning in the analogous non-tree interaction is

correct understanding of the non-tree interaction. Our main interest lies in computing the percentage of BI subject who lose – or, equivalently, fail at game-form recognition – as this metric shows how concerning the impact of failure in game-form recognition is.

To improve the test's reliability and robustness, the same test of game-form recognition is repeated in 27 different pairs of analogous tree and non-tree interactions. Each pair is treated as a separate experiment. The current replication crisis in social sciences (Maniadis et al. (2014), Camerer et al. (2016), Munafò et al. (2017), Camerer et al. (2018)) motivated the decision to replicate the same experiment in a variety of setups.

The pool of subjects consists of 4,788 people who collectively played 29,648 pairs of analogous interactions. Theory suggests that the percentage of BI subjects failing at game-form recognition should be zero. However, we found that this percentage varies from 12% to 65%. Even in the simplest interactions our subjects struggle with game-form recognition.

Documenting failure of game-form recognition is the first contribution of our paper. To position this contribution in a broader context, we make two observations. First, we recognize that game-form recognition is complementary to backward induction in a sense that a player's complete analysis of an interaction consists of understanding the interaction (game-form recognition) and solving it (backward induction). Consequently, the literature testing game-form recognition – that is, the literature that our paper belongs to – is complementary to the literature testing backward induction. Importantly, the former is small while the latter is vast; our paper attempts to fill the gap.

Second, we observe an analogy in interpreting empirical results. In the literature, the failure of backward induction is interpreted as indicating that backward induction is not a correct *model of behavior* because it does not capture what people choose in dynamic interactions. In the same spirit, the failure of game-form recognition tells us that a game-theoretic tree is not a correct *model of perception* since it does not depict how people perceive dynamic interactions.

The empirical evidence rejecting backward induction as a model of strategic behavior has inspired the development of alternative models (Camerer (2003), Crawford et al. (2013), Camerer and Ho (2015)). However, when it comes to players' perception, there is no alternative to the standard game-theoretic tree. Future research is needed to explore how players actually understand the games they participate in and provide accurate models of their perception. We hope that our paper inspires such theoretical developments.

Meanwhile, as another contribution of our paper, we attempt to answer a question that is

only natural to posit when confronted with the failure of both game-form recognition and backward induction: which task – game-form recognition or backward induction – is more challenging? To the best of our knowledge, this question has not been addressed in the literature to date. In the context of our experiment, we argue that subjects find game-form recognition to be more challenging than backward induction. Out of 27 pairs of tree/non-tree interactions, game-form recognition proves to be more difficult in 25 of them. This result highlights the importance of studying game-form recognition.

To collect the data, a group of professional app developers was recruited to create *Blues and Reds*, an app used to conduct mobile experiments. Here, "mobile" refers to the technology that the experiment relies on. *Blues and Reds* has been available for download since August 2017 by anyone in the world with an Android or iOS device and the ability to read in English, Spanish, traditional Chinese, or simplified Chinese. The app was designed to conduct various experiments. In Grabiszewski and Horenstein (2020a), we study the relationship between effort and skills. In Grabiszewski and Horenstein (2020b), we develop an empirical measure of complexity in dynamic games. In Grabiszewski and Horenstein (2020c), we offer a new method for profiling players in dynamic games.

The motivation behind conducting a mobile experiment is to exploit the widespread availability of mobile technology and smartphones in tandem with the increasing use of these devices. More people more frequently use mobile apps to make every day personal and professional decisions. However, this ubiquitous phenomenon has yet to be fully explored in academic research. Mobile experiments exploit this trend by reaching large and globally diversified samples and open the door to new frontiers in academic research by conducting experiments in the environment that is natural for a modern-day decision maker: on mobile devices.

From the methodological perspective, our paper belongs to the line of research that relies on nonstandard methods to gather data, like newspaper-based experiments (Bosch-Domènech et al. (2002)) and online experiments (Ariel Rubinstein's `gametheory.tau.ac.il`, Chen et al. (2014), Liu et al. (2014), and Chen and Konstan (2015)). When it comes to mobile experiments in economics, other than our projects based on *Blues and Reds*, we are only aware of Li et al. (2021) who study a variety of classical experimental games like the beauty contest.

In the context of individual decision-making, Cason and Plott (2014), Bull et al. (2019), and Drichoutis and Nayga (2020) report game-form misrecognition in the Becker-DeGroot-Marschak mechanism. Their subjects struggle with connecting their actions to their payoffs.

When it comes to game-form recognition in static games, the problem complementary to the one studied in this article, Devetag and Warglien (2008), Rydval et al. (2009), and Chou et al. (2009) design experiments employing the guessing game and observe that manipulating how a game is presented to the subjects changes their behavior.

Using the centipede game (Rosenthal (1981)), Cox and James (2012) and Crosetto and Mantovani (2018) focus on game-form recognition in finite dynamic games, the topic of this article. Crosetto and Mantovani (2018) alter presentation of payoffs to the subjects and find that making information less accessible makes subjects' choices deviate further from equilibrium. In our experiment, there is no room for misunderstanding the payoff structure, rather subjects fail at connecting their choices to their payoffs.

Cox and James (2012) employ an incomplete information version of the centipede game, while all games in our experiment are of complete information. Importantly, our subjects had unlimited time to make decisions while Cox and James (2012) impose a 10-second time limit for selecting actions. As argued in the literature (Reutskaja et al. (2011), De Paola and Gioia (2016), Spiliopoulos et al. (2018)), imposing time limits has a detrimental impact on the quality of subjects' choices. Therefore, what differentiates our paper from Cox and James (2012) is that we abstract from the impact of time-pressure on the selection of strategies. This allows us to isolate the failure of game-form recognition as the rationalization of subjects' observable behavior.

Additional contributions and factors differentiating our paper from the literature are the already mentioned replication success (our experiment consists of the same test conducted in 27 different settings) and the comparative analysis of difficulty of the two tasks that subjects accomplish (game-form recognition and backward induction).

Finally, the design of our experiment resolves a common concern present in the game-theoretic experiments on game-form recognition. As pointed out by Crosetto and Mantovani (2018), in such experiments, subjects deviate from theoretical predictions due to one of the following two reasons: (1) they misunderstand the game, or (2) they understand the game but believe that their opponents do not. Ideally, we would like to be able to attribute a subject's behavior solely to their correct/incorrect game-form recognition rather than to their belief about how well other players understand the interaction. After all, the goal of this paper is to test game-form recognition instead of studying what subjects believe about their opponents. In our experiment, it is the payoff structure and human subjects playing against the Artificial Intelligence that alleviate this concern.

The rest of this article is organized as follows. Section 2 describes the experimental de-

sign. Section 3 presents and discusses the results. Section 4 concludes. Appendix A contains the screenshots of all the tree and non-tree interactions from the experiment. Appendix B describes *Blues and Reds* from the gaming perspective and includes the experimental instructions. The relevant links for installing *Blues and Reds* are available at `www.bluesandreds.com`.

2 Experimental Design

2.1 Tree and non-tree interactions

Interactions in *Blues and Reds* are one of two types, tree and non-tree. A "tree interaction" is depicted as a classical game-theoretic tree, while a "non-tree interaction" is depicted in a more convoluted way. Figures 1 and 2 include screenshots from *Blues and Reds* with examples of tree and non-tree interactions, respectively.

[Figure 1 about here.]

[Figure 2 about here.]

Each interaction – tree or non-tree – is a finite dynamic game with complete and perfect information, and a no-tie zero-sum payoff structure. In each interaction, the subject plays against the Artificial Intelligence (AI) and can win. However, to win, the subject must follow the unique set of actions consistent with backward induction. If, at any round, the subject makes a mistake and deviates from that set, then she certainly loses as AI is designed to exploit the subject's mistakes. Note: AI is programmed to deterministically follow the backward induction algorithm.[1]

In *Blues and Reds*, subject and AI move the golden spherical object, called the RoboToken, across the blue (subject) and red (AI) bridges. Choices are made in turns with the subject moving first. At the last round, if the RoboToken lands on a blue node, then the subject wins; a red node signifies that the subject has lost.

To facilitate the discussion about how to play tree and non-tree interactions, we add labels for each node and reproduce Figure 1 in Figure 3, and Figure 2 in Figures 4 and 5.

[1]Subjects are not told that AI follows the backward induction algorithm. However, as we argue in Section 2.3, this is an irrelevant issue due to the structure of the games in our experiment.

[Figure 3 about here.]

[Figure 4 about here.]

[Figure 5 about here.]

In tree interactions, subjects can immediately backward induct to select their strategy. In Figure 3, the subject starts at node A and moves the RoboToken to either B_1 or B_2. If she selects B_1, then AI chooses among C_1, C_2, and C_3. If, however, she selects B_2, AI's options are C_4, C_5, and C_6. Once AI moves, the subject selects one of the D nodes. To win the tree interaction depicted in Figure 3, the subject must select B_2 at the first round. If the subject opts for B_1, then AI moves the RoboToken to C_1, which leaves the subject with only red nodes (D_1, D_2, and D_3) and results in a loss.

In non-tree interactions, subjects must begin with an understanding of the possible paths of play. The arrows help with this process. The blue arrows are located at rounds where the subject chooses an action. These indicate the maximum number of available moves and the direction of the move for the subject. The red arrows have the same role for AI.

In Figure 4, the subject begins at node A. The first right-pointing blue arrow with the number 2 indicates that the subject can select one of the two nodes located to the right of node A on the rail immediately above (B_1 or B_2). Choosing B_1 starts an animation that moves node A to the position that is immediately adjacent to node B_1. The blue bridge connects A and B_1, and the RoboToken crosses the bridge onto node B_1. Selecting B_2 moves node A across B_2, and the RoboToken lands on B_2.

No matter which B-node the subject chooses, AI follows the left-pointing red arrow with the number 3 and selects a node that is among the three nodes to the left of the current node that the RoboToken is on. This means that AI's options are C_1, C_2, and C_3 if the subject selected B_1 and, C_3, C_4, and C_5 if the subject chose B_2. After AI's move, the interaction ends with the subject choosing one of the D nodes in accordance with the right-pointing blue arrow with the number 3: D_1, D_2, or D_3 if the subject moves from C_1; D_3, D_4, or D_5 if the subject moves from C_2; D_5, D_6, or D_7 if the subject moves from C_3; D_7, D_8, or D_9 if the subject moves from C_4; and D_9, D_{10}, or D_{11} if the subject moves from C_5. To win the non-tree interaction represented in Figure 4, the subject must choose B_2 at the first round. Selecting B_1 will result in AI choosing C_2 which leaves the subject with a choice of nothing but a red node (D_3, D_4, and D_5).

To compare the tree interaction from Figure 3 with the non-tree interaction in Figure 4, we depict the latter in a tree form in Figure 5. We observe that the trees depicted in Figures 3 and 5 share an identical graph and fundamental game-theoretic properties mentioned at the beginning of this section. The difference between Figures 3 and 5 is the specific distribution of payoffs. However, for the purposes of our experiment, this difference is irrelevant: if the subject correctly backward inducts on a tree in Figure 3, then they will have no problem with backward induction on a tree in Figure 5 as the pertinent features (from the perspective of the backward induction algorithm) remain identical. This is precisely what makes the tree in Figure 1 and the non-tree in Figure 2 a pair of analogous interactions. In *Blues and Reds*, 58 interactions make up 29 pairs of analogous tree and non-tree interactions; two of these pairs make up the mandatory tutorial and are not part of our data.[2]

2.2 Structure of interactions

In each interaction (tree and non-tree), every node at a given round has the same number of branches. This symmetric structure allows us to label a pair of analogous interactions in a succinct way as $N_1.N_2.N_3.N_4.N_5.N_6$, where N_i is the number of actions at round i. For example, 2.3.3 is depicted in Figures 1 and 2. To simplify the notation, zeros are omitted. E.g., a 3-round interaction is labeled as $N_1.N_2.N_3$ instead of $N_1.N_2.N_3.0.0.0$.

The first column in Table 1 lists the 29 pairs of analogous tree and non-tree interactions which generated data for this article. The list excludes two pairs, 4.2 and 2.4.2, constituting the mandatory tutorial (see section 2.5 for more details about the tutorial).

2.3 Selecting backward-inducting subjects: empirical strategy

An integral part of the proposed test of game-form recognition is the selection of subjects who demonstrate the ability to backward induct. This test is conducted on tree interactions.

We claim that winning in a tree interaction indicates that a subject correctly backward inducted, while losing signifies a violation of the backward inducting hypothesis. For each tree interaction, we select subjects who win and call them the backward-inducting (BI) subjects. As discussed in section 2.1, in each tree interaction, winning is possible but requires following a unique set of actions as prescribed by backward induction. Consequently, a

[2]Additionally, 10 out of 27 pairs of analogous interactions that constitute our data have the same payoff distribution. All results presented in this paper hold if we just focus on those 10 interactions.

subject who correctly backward inducts wins with certainty.

However, it is only natural to ponder whether a losing subject could have also correctly backward inducted. After all, as it is well-known, testing backward induction is notoriously problematic due to the joint-hypothesis nature of the test implying that a subject who impeccably backward inducts might appear as if violating the hypothesis of backward induction.

First, there is the problem of social preferences. As an example, take the centipede game (Rosenthal (1981)). Under the assumption that players are selfish, backward induction predicts that the first player stops the game at the first node. Experimental evidence indicates that this is not the case (Krockow et al. (2016)). A typical explanation is the notion of altruism (McKelvey and Palfrey (1992)). A subject driven by altruism does not stop at the first node, a choice that should not be interpreted as a violation of backward induction.

Second, there is a problem of belief that the opponents are not fully rational (Palacios-Huerta and Volij (2009), Agranov et al. (2012), Alaoui and Penta (2016), Fehr and Huck (2016), Gill and Prowse (2016)). In the centipede game, if the first player assigns the probability high enough to their opponent being irrational, then they will not stop at the first node. Therefore, it would be a mistake to consider their choice as a violation of backward induction.

Social preferences or beliefs about other players' rationality change a subject's behavior. These phenomena generate noise that makes it impossible to test backward induction; consequently, elimination of this noise is critical. To address this problem, the tree interactions in *Blues and Reds* were designed in a way that minimizes, if not eliminates, the presence of these undesired phenomena and allows us to conclude that losing a tree interaction indicates behavior in disagreement with backward induction.

To address the problem of social preferences, we follow the approach advanced in the literature (Dufwenberg et al. (2010), Gneezy et al. (2010), Levitt et al. (2011)) and design all interactions in *Blues and Reds* as winner-takes-all games. However, even in such games it is possible for social preferences to be present; for instance, a parent loses on purpose in a tennis match played against their child. Hence, to err on the side of caution, *Blues and Reds* was designed as a human vs AI rather than a human vs human experiment. As Johnson et al. (2002) argue, playing against AI eliminates both the presence of and the subjects' beliefs in the presence of social preferences.

When it comes to subjects believing in their opponents' irrationality, it is the zero-sum structure of the payoffs that handles this problem (Levitt et al. (2011)). Even if a subject

assigns a non-zero probability to AI making a mistake, the subject has no reason to choose an action different from the one prescribed by backward induction.

To summarize, tree interactions allow us to test backward induction. Subjects who win a tree behave as if they correctly backward induct. Those who lose do not correctly implement the backward induction algorithm. In the second stage of our experiment, we focus solely on the BI subjects for testing game-form recognition.

2.4 Testing game-form recognition: empirical strategy

The hypothesis tested in this article is that the players in dynamic interactions correctly recognize the interactions they participate in. A player is said to correctly understand an interaction if they know the players, strategies, payoffs, connections among strategies and payoffs, and the sequences of choices.

Economic theory depicts dynamic interactions as game-theoretic trees. Naturally, it is only abstract players in academic papers and textbooks who think about dynamic interactions in terms of trees. For our subjects, as for all human decision-makers, we do not expect them to know about a game-theoretic tree or literally construct it. In fact, in many scenarios, drawing a tree is impractical (tic-tac-toe) or impossible (chess). To test whether subjects correctly recognize the interaction they play, ideally, we would like to directly observe how they actually perceive the interaction. Unfortunately, this is not the case. We do not observe the process of analyzing the interaction and how subjects perceive the interaction.

Hence, we resort to the "as if" approach, a standard methodology in economics. We do not study whether people apply game-theoretic tools and literally recognize interactions as trees. Rather, we test whether they behave as if a correctly constructed tree was their way of understanding an interaction. This is precisely in the same spirit that backward induction is tested in the literature. It is not hypothesized that people factually know about and apply the backward induction algorithm. Rather, it is tested whether people behave as if they were backward inducting; that is, whether they choose in accordance with predictions derived from the backward induction algorithm.

Our test exploits the perceptional difference between a tree interaction and the analogous non-tree interaction. To win a tree interaction, the subject must follow the unique set of actions consistent with backward induction. The only role that a tree interaction serves is that of a screening test selecting the BI subjects.

To win a non-tree interaction, it is necessary to correctly understand the interaction and, as in a tree interaction, choose in accordance with backward induction. If a subject succeeds at game-form recognition, then what remains for winning is to correctly apply backward induction on the model of non-tree interaction that the subject created in her mind. Hence, in order to test for game-form recognition, we look at the BI subjects because their behavior in a non-tree interaction can be interpreted in terms of game-form recognition. There are two possible scenarios.

First, if a BI subject wins the non-tree interaction, then we say that the subject behaves as if they correctly understand the interaction. In other words, the subject succeeds at game-form recognition.

Second, if a BI subject loses the non-tree interaction, then this indicates that there is a discrepancy between their perception and the correct understanding of the interaction. With no such discrepancy, the BI subject would win as confirmed by the fact that they just won in the analogous tree interaction.[3] While we do not know how a losing BI subject perceives a non-tree interaction in their mind, we do know that they failed at game-form recognition.

The main empirical result we are interested in is the percentage of BI subjects who win for each non-tree interaction. This percentage evaluates the hypothesis of game-form recognition. The lower the percentage of winning BI subjects, the more problematic this hypothesis is.

The proposed approach to test game-form recognition is similar to the experimental design in Levitt et al. (2011). Their objective is to determine how BI subjects behave in the centipede game. In the first step, they use the race game to identify BI subjects. In the second step, they look at the behavior of BI subjects in the centipede game. An important difference between this article and Levitt et al. (2011) is that we rely on analogous tree and non-tree interactions, a crucial feature of *Blues and Reds* allowing us to attribute the behavior of BI subjects to their correct/incorrect game-form recognition.

2.5 Experimental procedure

Recruitment of subjects took place via Google Play and the App Store; anyone who installs and plays *Blues and Reds* becomes a subject of the experiment. The app was promoted

[3]This also explains why we limit our attention to the BI subjects. A non-BI subject who also fails at game-form recognition might win in a non-tree interaction by incorrectly solving an incorrectly understood interaction. In this case, it would not be appropriate that this subject succeeded at game-form recognition.

with the help of AdWords, word of mouth, and media exposure (press in the United States, Poland, and Argentina).

Once the app is installed, subjects go through a mandatory tutorial consisting of 2 pairs of analogous tree and non-tree interactions. This mandatory tutorial supplies subjects with the experimental instructions. Subjects learn that they play in two-person interactions against AI in which they can either win or lose. Instructions indicate what must happen for the subjects to win/lose. Subjects are informed that their winning means that AI loses and that their losing indicates AI's winning. Last but not least, subjects learn how to make choices while playing tree and non-tree interactions in the app. Section B.4 in Appendix B contains a detailed description of the experimental instructions.

In the 54 interactions that constitute the core of the experiment, the subject can play each interaction only once, regardless of whether she wins or loses. This feature was introduced to motivate subjects to consciously think rather than mindlessly select their choices. The 27 pairs of analogous interactions are randomly allocated. In each pair, subjects play a tree interaction prior to playing the analogous non-tree interaction.

Subjects are not financially compensated for participating in the experiment or correctly solving the assigned tasks. Rather, the incentives are of non-monetary nature. As argued in the literature (Rubinstein (2016), Ding et al. (2018), Erkal et al. (2018)), such incentives are perfectly valid in experimental studies because they lead to an increase in the subjects' utility in the same way the monetary gains do.

Since the pool of subjects consists of self-selected people interested in mobile games about solving cognitive tasks, it is reasonable to posit that a subject's utility increases in the number of wins. In addition, there are in-game rewards: for winning a tree or non-tree interaction a subjects gains a star. Having more stars translates into a higher ranking in the app, which also has a positive impact on the subject's utility. Finally, there is evidence that virtual items do indeed have monetary value (Drummond et al. (2020)). We refer to Appendix B for more details about the in-game reward system.

3 Empirical Analysis

3.1 Data

Data was collected from August 15, 2017 to February 6, 2018. For each interaction and each subject, *Blues and Reds* records (1) whether the subject wins or loses, and (2) their response times (RT) measured in seconds at each round. Given the objective of our experiment, we remove all observations in which a subject played a tree interaction but did not participate in the analogous non-tree interaction. In each interaction, observations with a total RT above the 90th percentile were removed.[4]

Each observation consists of data for a given subject and a given pair of analogous tree and non-tree interactions. The final data set consists of 29,648 observations generated by 4,788 subjects. The average number of pairs of interactions played by a subject is 6.19 (with a standard deviation of 6.54). In terms of the geographic distribution of the sample, subjects represent 137 countries. The most popular countries in the sample are Mexico (13%), India (11%), the United States (9%), Argentina (8%), Brazil (5%), Colombia (5%), and Poland (4%).

3.2 Overview of empirical exercises

As discussed in section 2.3, a BI subject winning (losing) a non-tree interaction indicates a success (failure) of game-form recognition. For each non-tree interaction, we compute the percentage of winning BI subjects. The non-negligible failure of game-form recognition across non-tree interactions is the main result of this paper and is presented in section 3.3.

Next, we check whether subjects exert more effort when dealing with non-tree interactions than the analogous tree interactions. This is an expected result as non-trees call for game-form recognition. As it is standard in the literature, we use total RT as a measure of cognitive effort (Rubinstein (2006), Clithero (2018), Spiliopoulos and Ortmann (2017), Coricelli et al. (2020)). We compare the average total RT between a tree and the analogous non-tree interaction. The results are presented in section 3.4.

Finally, we ask whether understanding a game is more difficult than solving it. Our conjecture is based on the analysis of conditional probabilities. The results are presented in section

[4] All results hold if we remove data with total RT above the 95th or 99th percentile. Given our large sample, we decided to present the results removing data above 90th percentile to err on the side of caution regarding the existence of outliers.

3.5.

3.3 Testing game-form recognition: results

Table 1 presents the main results from the experiment. Column A lists the 27 different tree and non-tree interactions. For a given interaction, Column B depicts the number of subjects who played both tree and non-tree versions of the interaction.

Column C includes the percentage of subjects who won the tree interaction. Recall that winning a tree is attributed to correct backward induction; hence, the winners are called BI subjects. Results in Column C confirm what is already known in the literature: in general, people struggle with backward induction. The percentage of subjects who behave consistently with backward induction ranges from as little as 43.64% to 98.47%.

Column D provides the number of BI subjects who played a given non-tree interaction. Column E includes the key results of the experiment: the percentage of BI subjects who won the non-tree interaction. Recall that a BI subject's winning a non-tree interaction indicates a success of game-form recognition.

[Table 1 about here.]

In every non-tree interaction, there is a significant proportion of BI subjects who lose. This is true even for the simplest interactions – those with over 90% of subjects backward inducting in the analogous tree interaction. The percentage of subjects who fail at game-form recognition ranges from 12.22% in 2.2.3 to 64.81% in 3.2.2.2.2. While Myerson (1991) argues that "whatever model of the game we may study, we must assume that the players know this model," the results in our experiment suggest caution against making this assumption.

Some readers might ponder whether framing (Tversky and Kahneman (1981)) is a phenomenon present in our experiment. Undoubtedly, while framing matters in the context of dynamic games (Camerer et al. (1993)), whether it also matters in our experiment depends on how "framing" is defined. On the one hand, framing might be understood broadly as a problem of perception: choices depend on the particular presentation of a problem. In this case, clearly, framing is relevant for our experiment.

On the other hand, Tversky and Kahneman (1981) define framing more narrowly as "predictable shifts of preference when the same problem is framed in different ways." Chou et al. (2009) and Cason and Plott (2014) interpret framing in a similar spirit. From this

perspective, there is no issue of framing in our experiment because in *Blues and Reds* there is no shift of preferences as each interaction ends with the subject either winning or losing.

3.4 Effort in tree and non-tree interactions

By design, a non-tree interaction is more difficult to solve than the analogous tree interaction. The former requires more effort when going through the process of game-form recognition. To verify whether, as expected, subjects actually spend more effort on non-trees, we conduct the following exercise.

A subject's cognitive effort is measured as the total RT they spend solving an interaction (Rubinstein (2006), Clithero (2018) Spiliopoulos and Ortmann (2017), Coricelli et al. (2020)). We compute the average total response times for each tree and non-tree. The results are presented in Table 2. Column A lists the 27 tree and non-tree interactions.

Column B presents data computed for the whole sample. In every tree/non-tree pair, the average total response time is larger in the non-tree interaction. In each case, the difference is statistically significant at less than the 1% level.

As a robustness check, we study the differences in average total response times for various subsamples. Column C depicts the results for losers: those who lost a tree interaction and those who lost a non-tree interaction. Column D presents the data for tree winners and non-tree winners. Finally, Column E consists of averages in a tree interaction and non-tree interaction computed for subjects who won both the tree and the analogous non-tree.

For every tree/non-tree pair and every column, the average total response time is larger for the non-tree interaction. This confirms that subjects do spend more effort analyzing and solving non-tree interactions than the analogous tree interactions.

[Table 2 about here.]

3.5 Game-form recognition versus backward induction: what is more difficult?

Using the results presented in Table 1, we analyze whether the task of game-form recognition is more difficult than the task of backward induction. Let $P(BI)$ be the probability of a subject correctly backward inducting in a tree interaction. For example, $P(BI) = 96.43\%$

in the tree 2.2. Let $P(GFR)$ stand for the probability of a subject succeeding at game-form recognition in a non-tree interaction. The goal is to determine whether, for a given pair of analogous tree and non-tree interactions, $P(GFR)$ is larger than $P(BI)$. If this is the case, then we conclude that game-form recognition is more difficult than backward induction.

The main challenge is that we do not directly observe $P(GFR)$. Rather, our data in Table 1 includes $P(GFR|BI)$, the conditional probability of a BI subject correctly understanding a non-tree interaction.

To study $P(GFR)$, we rely on the Bayes' formula, $P(GFR) = P(GFR|BI) \times \frac{P(BI)}{P(BI|GFR)}$. We posit that $P(BI)$ at most equals $P(BI|GFR)$ as those who succeed at game-form recognition should not be worse at backward induction compared to the whole population. Note, however, that this is not a hypothesis we can test in our data; rather, we find this to be a plausible assumption.

With $\frac{P(BI)}{P(BI|GFR)} \leq 1$, the upper bound for $P(GFR)$ is $P(GFR|BI)$. Hence, if for a given pair of analogous tree and non-tree interactions $P(GFR|BI) < P(BI)$, then it is the case that $P(GFR) < P(BI)$. That is, backward induction on a tree is simpler than game-form recognition on the analogous non-tree.

Analyzing Column C depicting $P(BI)$ and Column E depicting $P(GFR|BI)$ in Table 1, we observe that $P(GFR) < P(BI)$ in 25 cases. In 24 out of the 25 cases, the difference $P(GFR) - P(BI)$ is smaller than zero and statistically significant at the 1% level or less. For 4.2.2.2.2, the difference is positive but not statistically different from zero. It is only for 2.3.2.2 that the difference $P(GFR) - P(BI))$ is larger than zero and statistically significant, but only at the 10% level.[5]

We conclude that, in our experiment, game-form recognition is more challenging than backward induction. This conclusion is in line with Ackoff (1974) who argues that "we fail more often because we solve the wrong problem than because we get the wrong solution to the right problem" and Albert Einstein who remarked that "if I had an hour to solve a problem, I'd spend 55 minutes thinking about the problem and five minutes thinking about solutions."

It is important to stress that the finding of backward induction being easier than game-form recognition might be a result driven entirely by the specific design of our experiment. With a different visual presentation of non-tree interactions, we might observe backward induction to be more difficult. For that reason, the results presented in this section should be treated with caution. Nevertheless, the result is informative and highlights the substantial impact

[5]For calculating the significance level of the difference between $P(BI)$ and $P(GRF)$ we compute the Welch's t-test.

that game-form recognition can have on the behavior of subjects that can correctly backward induct.

4 Conclusions

The main question this article addresses is whether people succeed at game-form recognition in dynamic interactions representable as trees and solvable by backward induction. The test consists of two types of analogous interactions, tree and non-tree, and two stages.

First, the tree interaction serves as a screening test to select the BI subjects. Second, the BI subjects are studied in the analogous non-tree interaction. Losing and winning in the second stage is attributed to incorrect and correct game-form recognition, respectively.

There are 27 analogous pairs of interactions in the experiment. The sample consists of 4,788 subjects representing 137 countries. The percentage of BI subjects who fail at game-form recognition ranges from 12% to 65%. This indicates that people do not necessarily understand the dynamic interactions they play. Finally, in the context of our experiment, subjects are more likely to correctly backward induct than correctly understand a game.

To collect data, we conduct a mobile experiment. For social scientists, the smartphone is an innovation like no other as it offers a unique opportunity to conduct large-scale global experiments. With this motivation in mind, the mobile app *Blues and Reds* was commissioned and became available for free on both iOS and Android devices in August 2017. As a proof of concept, *Blues and Reds* shows that mobile technology offers exciting opportunities for academic research.

References

Ackoff, R. L. (1974): *Redesigning the Future: A Systems Approach to Societal Problems*, New York, NY: John Wiley & Sons.

Agranov, M., E. Potamites, A. Schotter, and C. Tergiman (2012): "Beliefs and endogenous cognitive levels: An experimental study," *Games and Economic Behavior*, 75, 449–463.

Alaoui, L. and A. Penta (2016): "Endogenous Depth of Reasoning," *Review of Economic Studies*, 83, 1297–1333.

Bosch-Domènech, A., J. G. Montalvo, R. Nagel, and A. Satorra (2002): "One, Two, (Three), Infinity, ... : Newspaper and Lab Beauty-Contest Experiments," *American Economic Review*, 92, 1687–1701.

Bull, C., P. Courty, M. Doyon, and D. Rondeau (2019): "Failure of the Becker-DeGroot-Marschak mechanism in inexperienced subjects: New tests of the game form misconception hypothesis," *Journal of Economic Behavior & Organization*, 159, 235–253.

Camerer, C. F., ed. (2003): *Behavioral Game Theory: Experiments in Strategic Interaction*, Princeton, NJ: Princeton University Press.

Camerer, C. F., A. Dreber, E. Forsell, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, J. Almenberg, A. Altmejd, T. Chan, E. Heikensten, F. Holzmeister, T. Imai, S. Isaksson, G. Nave, T. Pfeiffer, M. Razen, and H. Wu (2016): "Evaluating replicability of laboratory experiments in economics," *Science*, 351, 1433–1436.

Camerer, C. F., A. Dreber, F. Holzmeister, T.-H. Ho, J. Huber, M. Johannesson, M. Kirchler, G. Nave, B. A. Nosek, T. Pfeiffer, A. Altmejd, N. Buttrick, T. Chan, Y. Chen, E. Forsell, A. Gampa, E. Heikensten, L. Hummer, T. Imai, S. Isaksson, D. Manfredi, J. Rose, E.-J. Wagenmakers, and H. Wu (2018): "Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015," *Nature Human Behaviour*, 2, 637–644.

Camerer, C. F. and T.-H. Ho (2015): "Behavioral Game Theory Experiments and Modeling," in *Handbook of Game Theory with Economic Applications*, ed. by H. P. Young and S. Zamir, Amsterdam, The Netherlands: North Holland, vol. 4, 517–573.

Camerer, C. F., E. J. Johnson, T. Raymon, and S. Sen (1993): "Cognition and Framing in Sequential Bargaining for Gains and Losses," in *Frontiers of game theory*, ed. by K. Binmore, A. Kirman, and P. Tani, Cambridge, MA: MIT Press, 27–47.

Cason, T. N. and C. R. Plott (2014): "Misconceptions and Game Form Recognition: Challenges to Theories of Revealed Preference and Framing," *Journal of Political Economy*, 122, 1235–1270.

Chen, Y., G. Y. Jeon, and Y.-M. Kim (2014): "A day without a search engine: an experimental study of online and offline searches," *Experimental Economics*, 17, 512–536.

Chen, Y. and J. Konstan (2015): "Online field experiments: a selective survey of methods," *Journal of the Economic Science Association*, 1, 29–42.

Chou, E., M. McConnell, R. Nagel, and C. Plott (2009): "The Control of Game Form Recognition in Experiments: Understanding Dominant Strategy Failures in a Simple Two Person "Guessing" Game," *Experimental Economics*, 12, 159–179.

Clithero, J. A. (2018): "Response Times in Economics: Looking Through the Lens of Sequential Sampling Models," *Journal of Economic Psychology*, 69, 61–86.

Coricelli, G., L. Polonio, and A. Vostroknutov (2020): "The Process of Choice in Games," in *Handbook of Experimental Game Theory*, ed. by C. M. Capra, R. T. A. Croson, M. L. Rigdon, and T. S. Rosenblat, Edward Elgar Publishing.

Cox, J. C. and D. James (2012): "Clocks and Trees: Isomorphic Dutch Auctions and Centipede Games," *Econometrica*, 80, 883–903.

Crawford, V. P., M. A. Costa-Gomes, and N. Iriberri (2013): "Structural Models of Nonequilibrium Strategic Thinking: Theory, Evidence, and Applications," *Journal of Economic Literature*, 51, 1–15.

Crosetto, P. and M. Mantovani (2018): "Representation effects in the centipede game," *PLoS ONE*, 13, e0204422.

De Paola, M. and F. Gioia (2016): "Who performs better under time pressure? Results from a field experiment," *Journal of Economic Psychology*, 53, 37–53.

Devetag, G. and M. Warglien (2008): "Playing the Wrong Game: An Experimental Analysis of Relational Complexity and Strategic Misrepresentation," *Games and Economic Behavior*, 62, 364–382.

Ding, S., V. Lugovskyy, D. Puzzello, S. Tucker, and A. Williams (2018): "Cash versus extra-credit incentives in experimental asset markets," *Journal of Economic Behavior & Organization*, 150, 19–27.

Drichoutis, A. and R. Nayga (2020): "Game form recognition in preference elicitation, cognitive abilities and cognitive load," *MPRA Paper No. 97980*.

Drummond, A., J. D. Sauer, L. C. Hall, D. Zendle, and M. R. Loudon (2020): "Why loot boxes could be regulated as gambling," *Nature Human Behaviour*, 4, 986–988.

Dufwenberg, M., R. Sundaram, and D. J. Butler (2010): "Epiphany in the Game of 21," *Journal of Economic Behavior & Organization*, 75, 132–143.

Erkal, N., L. Gangadharan, and B. H. Koh (2018): "Monetary and non-monetary incentives in real-effort tournaments," *European Economic Review*, 101, 528–545.

Fehr, D. and S. Huck (2016): "Who Knows It is a Game? On Strategic Awareness and Cognitive Ability," *Experimental Economics*, 19, 713–726.

Gill, D. and V. Prowse (2016): "Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A Level-k Analysis," *Journal of Political Economy*, 124, 1619–1676.

Gneezy, U., A. Rustichini, and A. Vostroknutov (2010): "Experience and insight in the Race game," *Journal of Economic Behavior & Organization*, 75, 144–155.

Grabiszewski, K. and A. Horenstein (2020a): "Effort is not a monotonic function of skills: results from a global mobile experiment," *Journal of Economic Behavior & Organization*, 176.

——— (2020b): "Measuring tree complexity with response times," *working paper*.

——— (2020c): "Profiling Players in Dynamic Games: A Mobile Experiment," *working paper*.

Johnson, E. J., C. Camerer, S. Sen, and T. Rymon (2002): "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining," *Journal of Economic Theory*, 104, 16–47.

Krockow, E. M., A. M. Colman, and B. D. Pulford (2016): "Cooperation in repeated interactions: A systematic review of Centipede game experiments, 1992 – 2016," *European Review of Social Psychology*, 27, 231–282.

Levitt, S. D., J. A. List, and S. E. Sadoff (2011): "Checkmate: Exploring Backward Induction among Chess Players," *American Economic Review*, 101, 975–990.

Li, Z., P.-H. Lin, S.-Y. Kong, D. Wang, and J. Duffy (2021): "Conducting large, repeated, multi-game economic experiments using mobile platforms," *PLoS ONE*, 16, e0250668.

Liu, T. X., J. Yang, L. A. Adamic, and Y. Chen (2014): "Crowdsourcing with All-Pay Auctions: A Field Experiment on Taskcn," *Management Science*, 60, 2020–2037.

Maniadis, Z., F. Tufano, and J. A. List (2014): "One Swallow Doesn't Make a Summer: New Evidence on Anchoring Effects," *American Economic Review*, 104, 277–290.

McKelvey, R. D. and T. R. Palfrey (1992): "An Experimental Study of the Centipede Game," *Econometrica*, 60, 803–836.

Munafò, M. R., B. A. Nosek, D. V. M. Bishop, K. S. Button, C. D. Chambers, N. P. du Sert, U. Simonsohn, E.-J. Wagenmakers, J. J. Ware, and J. P. Ioannidis (2017): "A manifesto for reproducible science," *Nature Human Behaviour*, 1, 0021.

Myerson, R. B. (1991): *Game Theory*, Cambridge, MA: Harvard University Press.

Palacios-Huerta, I. and O. Volij (2009): "Field Centipedes," *American Economic Review*, 9, 1619–1635.

Reutskaja, E., R. Nagel, C. F. Camerer, and A. Rangel (2011): "Search Dynamics in Consumer Choice under Time Pressure: An Eye-Tracking Study," *American Economic Review*, 101, 900–926.

Rosenthal, R. W. (1981): "Games of perfect information, predatory pricing and the chain-store paradox," *Journal of Economic Theory*, 25, 92–100.

Rubinstein, A. (2006): "Dilemmas of an Economic Theorist," *Econometrica*, 74, 865–883.

——— (2016): "A Typology of Players: Between Instinctive and Contemplative," *Quarterly Journal of Economics*, 131, 859–890.

Rydval, O., A. Ortmann, and M. Ostatnicky (2009): "Three Very Simple Games and What It Takes to Solve Them," *Journal of Economic Behavior & Organization*, 72, 589–601.

Spiliopoulos, L. and A. Ortmann (2017): "The BCD of Response Time Analysis in Experimental Economics," *Experimental Economics*, 47, 1–55.

Spiliopoulos, L., A. Ortmann, and L. Zhang (2018): "Complexity, attention, and choice in games under time constraints: A process analysis," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44, 1609–1640.

Tversky, A. and D. Kahneman (1981): "The Framing of decisions and the psychology of choice," *Science*, 211, 453–458.

Figure 1: Tree interaction. A screenshot from *Blues and Reds* with a tree interaction.

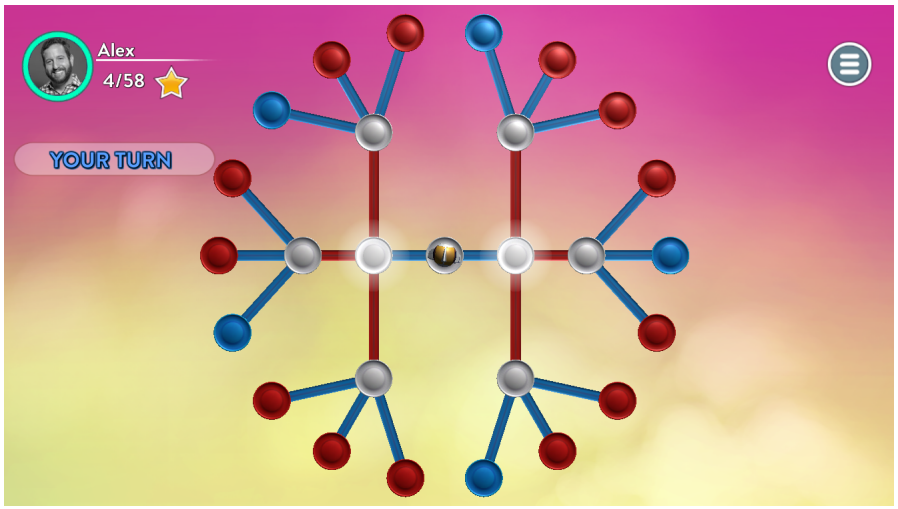


Figure 2: Non-tree interaction. A screenshot from *Blues and Reds* with a non-tree interaction.

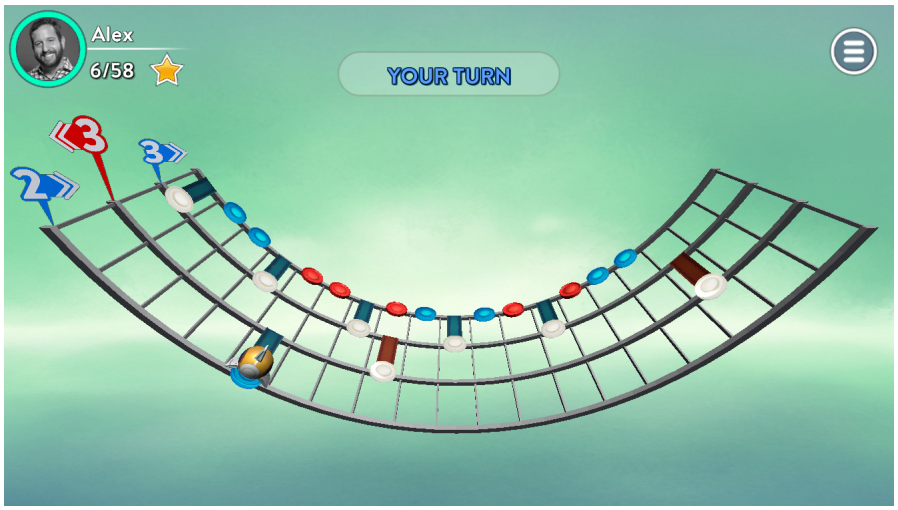


Figure 3: Tree interaction with node labels.

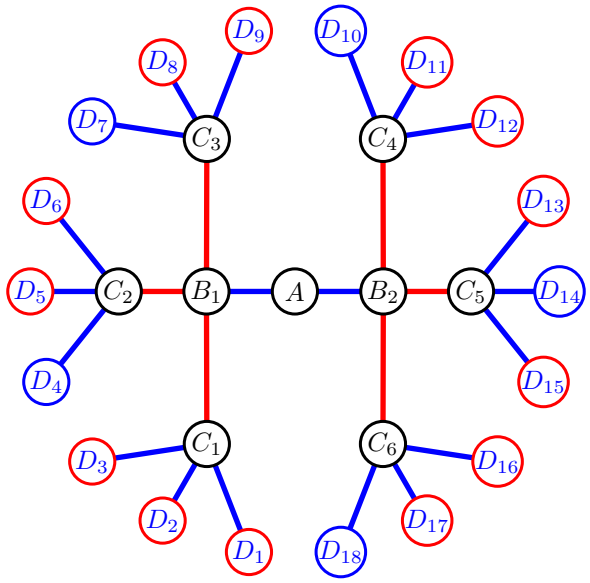


Figure 4: Non-tree interaction with node labels.

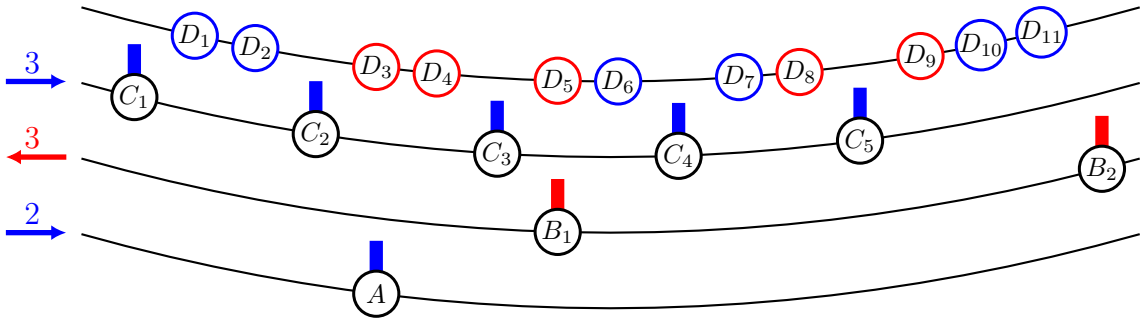


Figure 5: Non-tree interaction depicted as a tree with node labels.

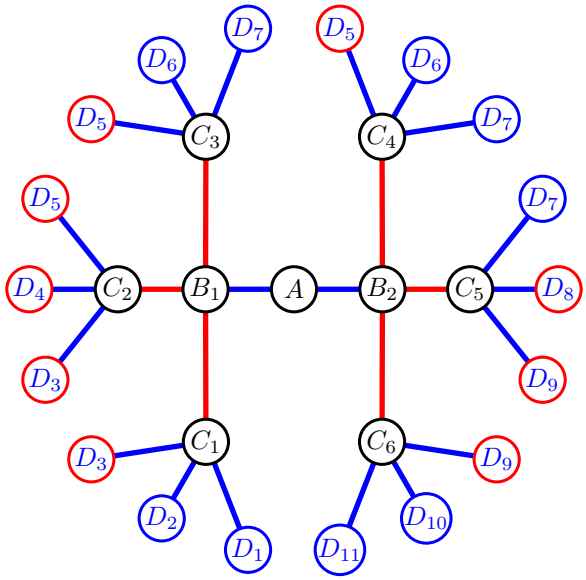


Table 1: Main results.

A	B	C	D	E
	all subjects		BI subjects	
interaction	N	% won tree	N	% won non-tree
2.2	1,261	96.43%	1,216	78.95%
2.3	1,244	98.47%	1,225	81.88%
3.2	1,261	97.22%	1,226	64.27%
3.3	1,214	94.40%	1,146	85.60%
2.4	1,177	97.71%	1,150	81.30%
2.2.2	1,108	95.13%	1,054	86.05%
2.3.2	1,099	93.54%	1,028	80.64%
2.2.3	1,137	95.69%	1,088	87.78%
3.2.2	1,116	93.55%	1,044	65.61%
2.3.3	1,176	94.73%	1,114	79.89%
3.3.2	1,131	91.87%	1,039	64.29%
3.2.3	1,085	93.36%	1,013	77.00%
3.3.3	1,117	91.41%	1,021	63.17%
4.2.2	1,121	91.35%	1,024	64.16%
2.2.2.2	1,131	68.52%	775	55.74%
2.3.2.2	1,128	59.57%	672	63.99%
2.2.3.2	1,111	79.48%	883	55.15%
2.2.2.3	1,100	81.55%	897	59.09%
3.2.2.2	1,053	71.80%	756	48.68%
2.4.2.2	1,167	82.35%	961	63.27%
2.2.4.2	1,095	75.53%	827	49.94%
2.2.2.4	1,100	84.27%	927	53.61%
4.2.2.2	1,106	71.16%	787	43.46%
2.2.2.2.2	855	68.89%	589	49.75%
3.2.2.2.2	856	65.07%	557	35.19%
4.2.2.2.2	803	44.71%	359	45.68%
2.2.2.2.2.2	896	43.64%	391	39.39%

Table 2: Average total response time in tree and non-tree interactions.

A	B		C		D		E	
	all subjects		losers		winners		won tree & non-tree	
interaction	tree	non-tree	tree	non-tree	tree	non-tree	tree	non-tree
2.2	8.5	12.6	8.8	12.8	8.5	12.5	8.4	12.5
2.3	8.8	11.9	10.9	11.9	8.8	11.9	8.6	11.9
3.2	9.0	17.7	10.5	19.4	9.0	16.7	8.7	16.7
3.3	9.8	14.0	10.0	14.3	9.8	13.9	9.7	13.9
2.4	8.8	12.5	9.8	12.2	8.8	12.6	8.8	12.6
2.2.2	18.0	29.4	21.6	35.9	17.8	28.3	17.7	28.5
2.3.2	19.6	39.3	23.4	44.7	19.4	38.0	19.2	38.5
2.2.3	20.3	29.4	23.5	34.4	20.1	28.6	19.9	28.6
3.2.2	19.7	43.2	22.3	48.8	19.5	39.9	18.8	40.0
2.3.3	21.2	38.4	26.0	46.5	20.9	36.2	20.7	36.5
3.3.2	21.8	45.1	24.4	48.9	21.6	42.8	21.6	43.4
3.2.3	20.6	39.9	22.8	46.5	20.4	37.8	20.2	38.0
3.3.3	24.0	47.5	26.8	49.8	23.8	46.0	23.3	47.1
4.2.2	21.0	44.0	24.9	45.3	20.6	43.2	20.3	43.7
2.2.2.2	27.7	40.8	30.9	37.0	26.2	45.7	25.5	45.5
2.3.2.2	33.4	51.9	34.8	42.8	32.4	61.8	32.6	63.1
2.2.3.2	32.2	52.5	33.4	48.8	31.9	56.8	31.6	57.1
2.2.2.3	31.1	41.2	32.8	37.4	30.7	44.7	29.9	45.3
3.2.2.2	35.4	50.6	34.5	44.4	35.8	60.1	34.7	62.0
2.4.2.2	33.0	56.1	34.1	48.6	32.7	61.3	33.3	61.3
2.2.4.2	37.1	50.5	39.8	43.4	36.2	60.5	37.2	61.6
2.2.2.4	30.4	47.3	29.3	41.5	30.6	53.5	30.5	54.9
4.2.2.2	35.3	54.9	34.9	46.3	35.4	71.5	35.5	73.1
2.2.2.2.2	57.2	87.4	58.4	84.4	56.6	91.2	59.3	97.7
3.2.2.2.2	62.3	93.5	56.1	84.6	65.6	114.8	73.6	128.2
4.2.2.2.2	83.0	110.7	69.9	101.0	99.1	129.2	117.2	169.2
2.2.2.2.2.2	71.2	84.0	62.6	74.2	82.3	114.1	102.2	126.8

A Appendix A: Screenshots of interactions in *Blues and Reds*

Figure 1: Tree 2.2

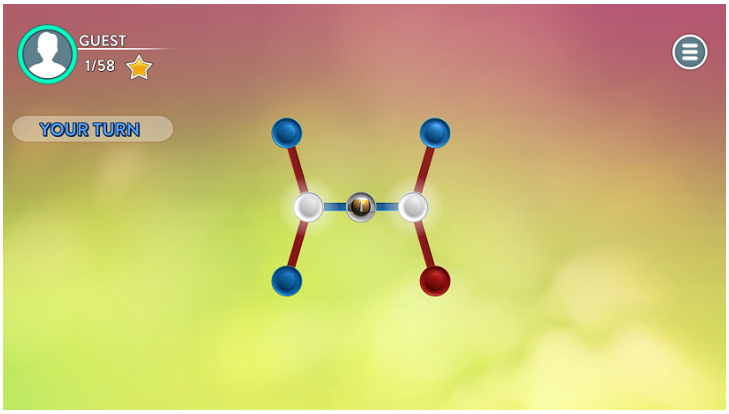


Figure 2: Non-tree 2.2

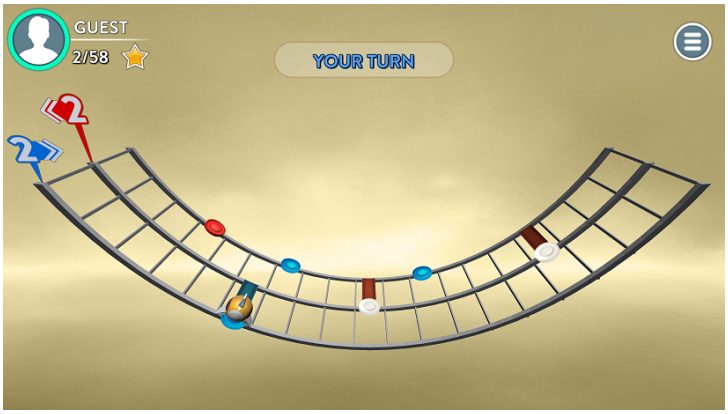


Figure 3: Tree 2.3

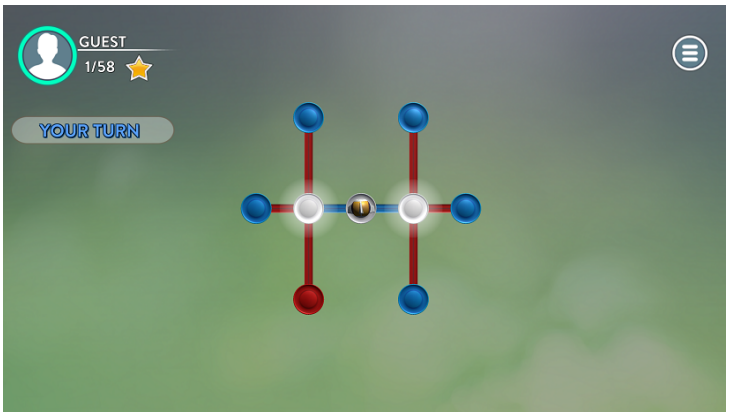


Figure 4: Non-tree 2.3

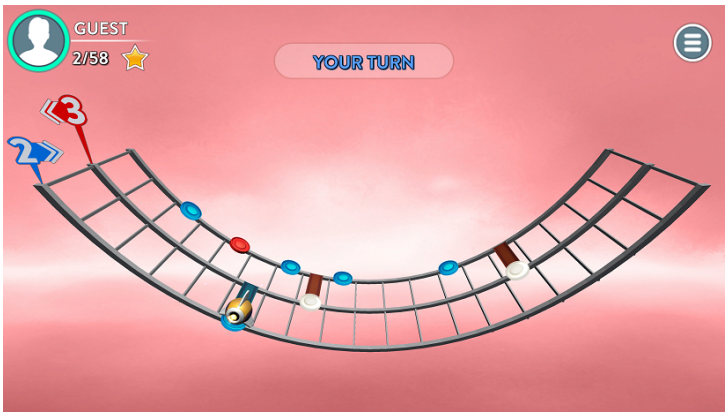


Figure 5: Tree 2.4

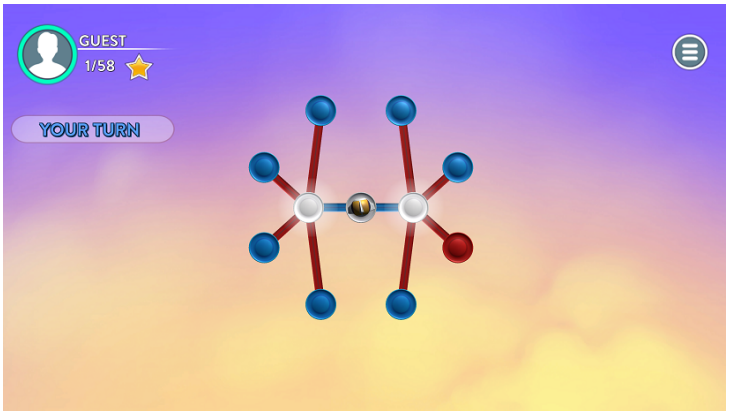


Figure 6: Non-tree 2.4

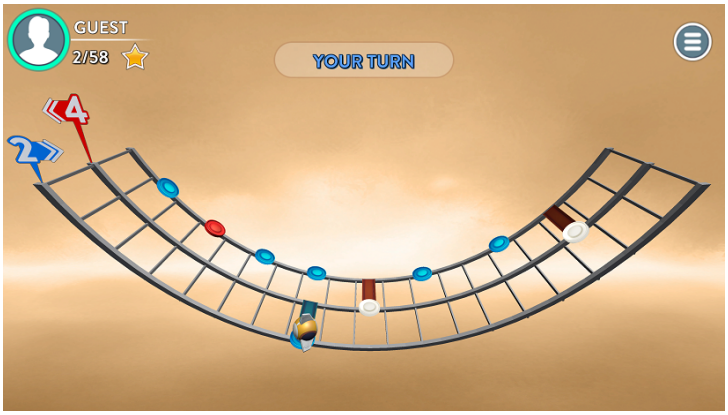


Figure 7: Tree 3.2

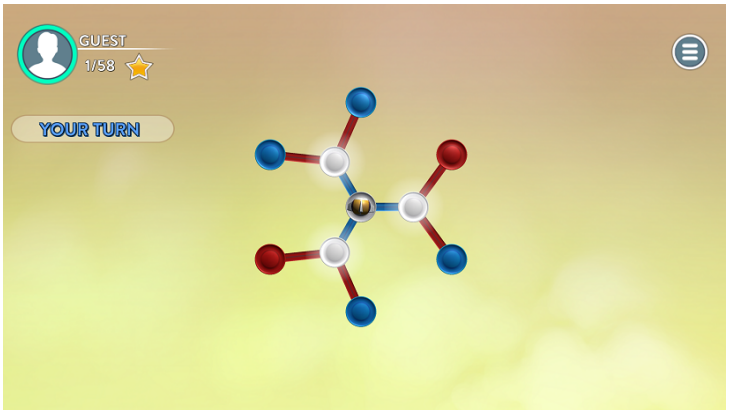


Figure 8: Non-tree 3.2

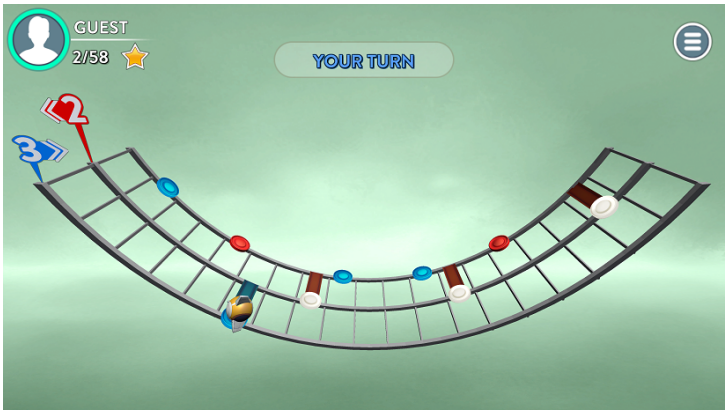


Figure 9: Tree 3.3

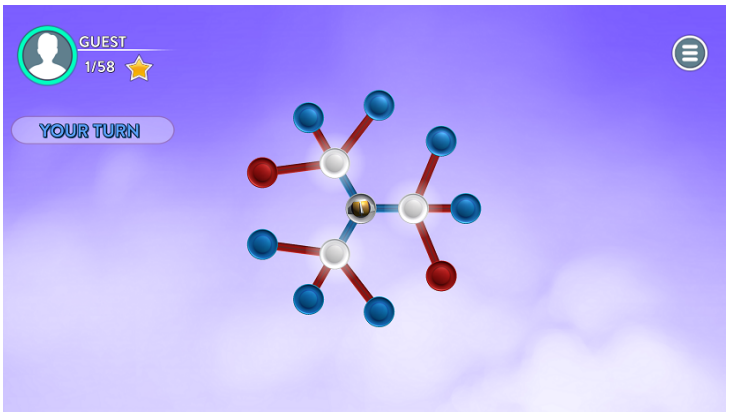


Figure 10: Non-tree 3.3

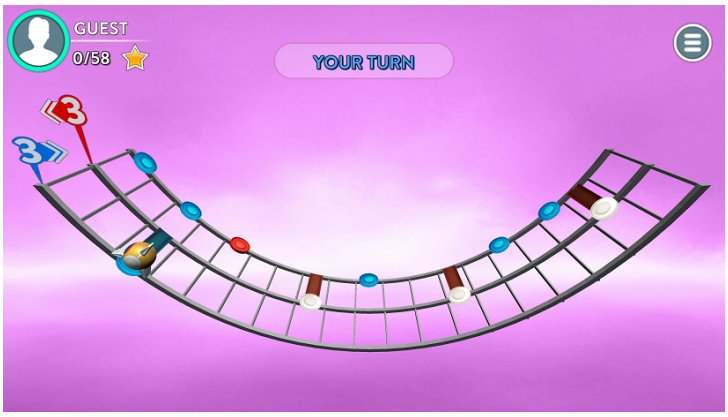


Figure 11: Tree 2.2.2

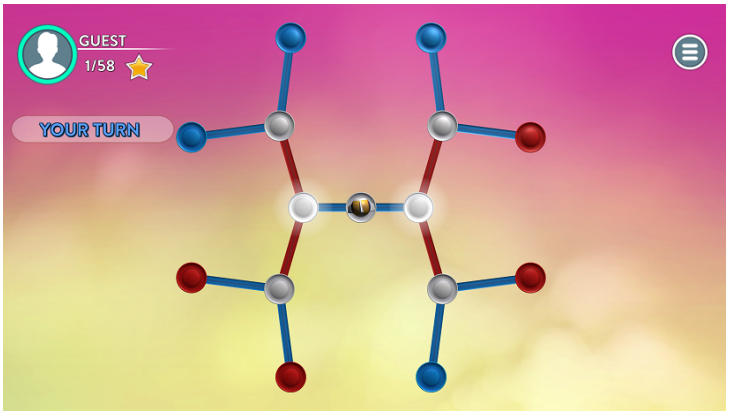


Figure 12: Non-tree 2.2.2

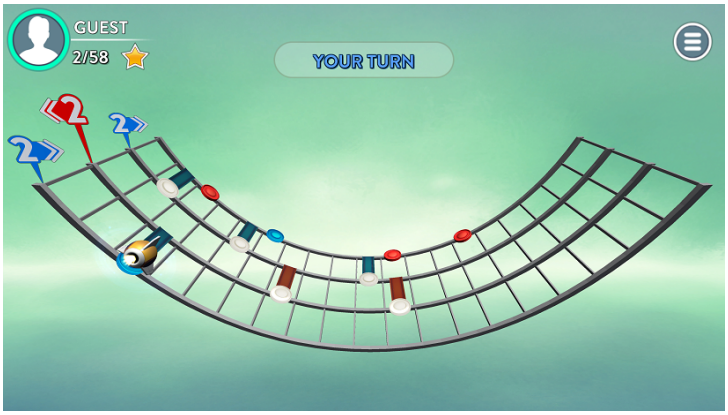


Figure 13: Tree 2.2.3

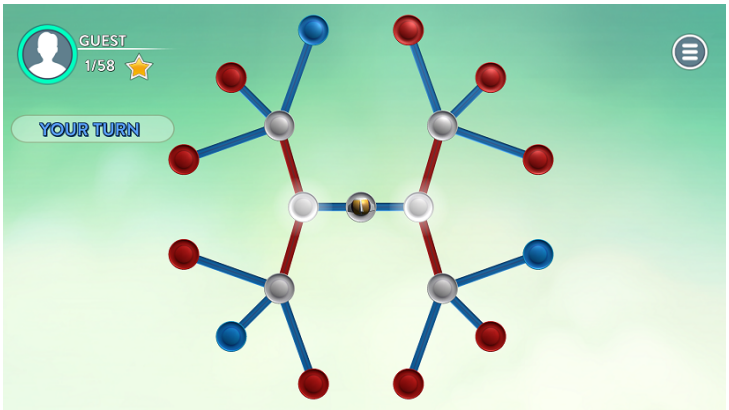


Figure 14: Non-tree 2.2.3

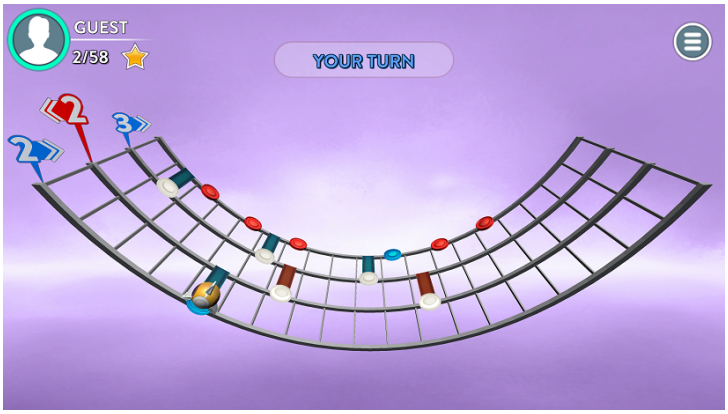


Figure 15: Tree 2.3.2

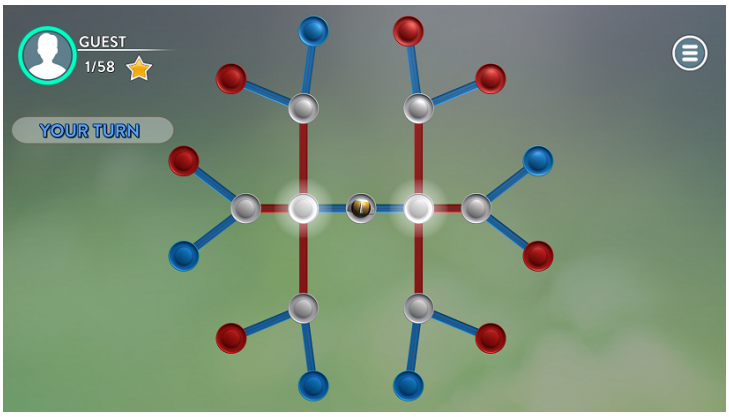


Figure 16: Non-tree 2.3.2

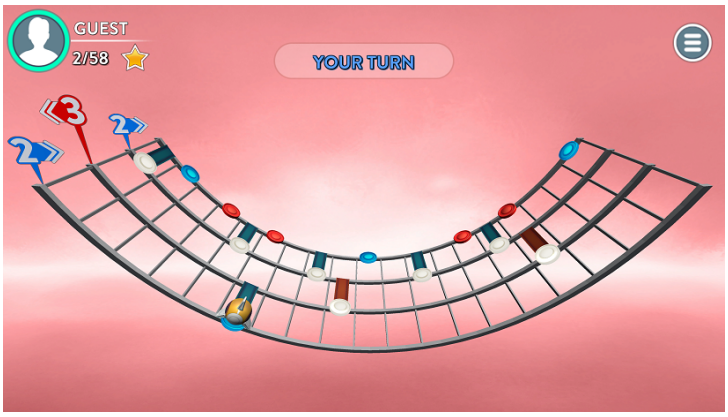


Figure 17: Tree 2.3.3

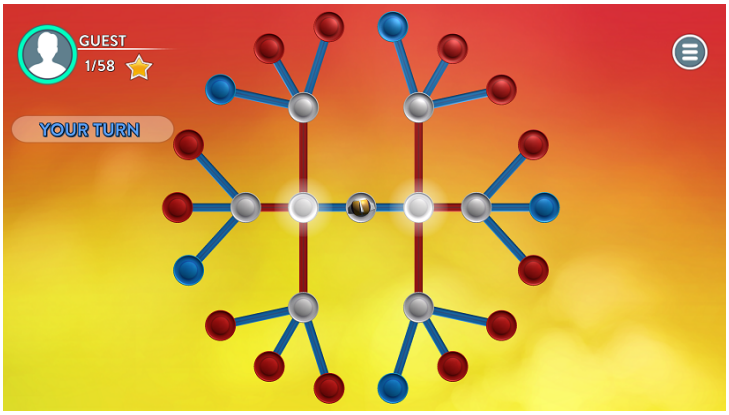


Figure 18: Non-tree 2.3.3

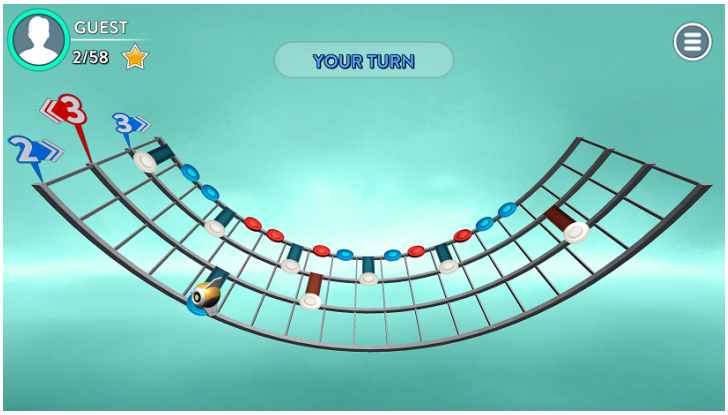


Figure 19: Tree 3.2.2

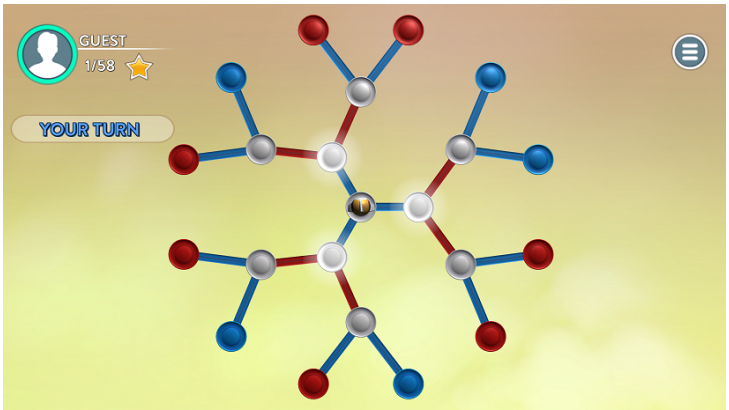


Figure 20: Non-tree 3.2.2

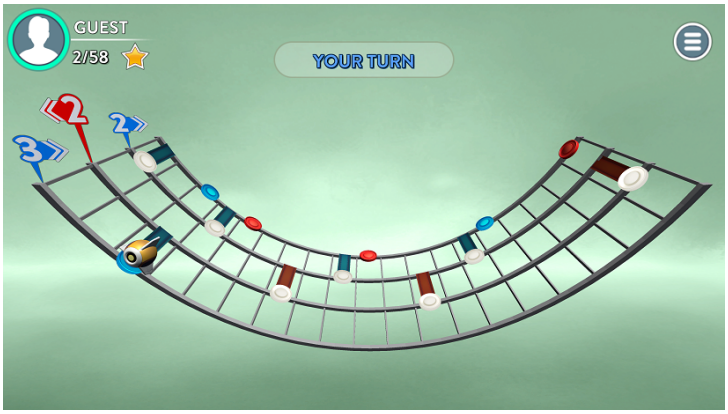


Figure 21: Tree 3.2.3

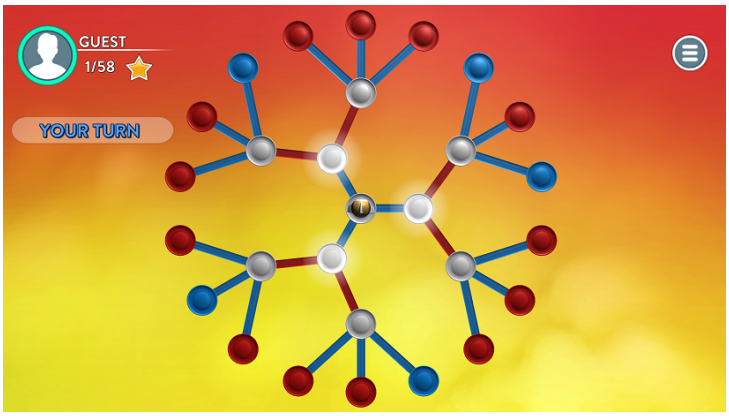


Figure 22: Non-tree 3.2.3

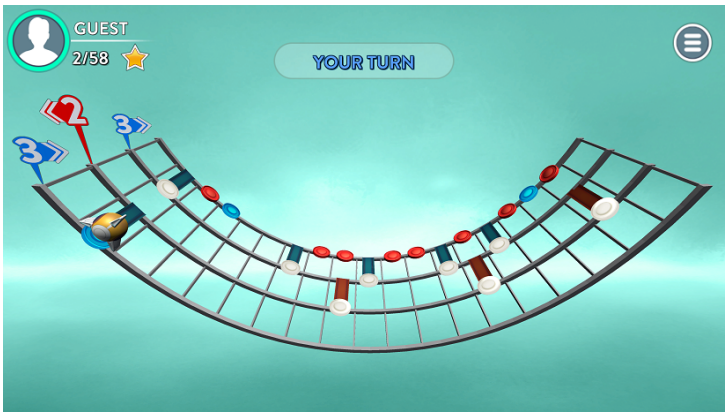


Figure 23: Tree 3.3.2

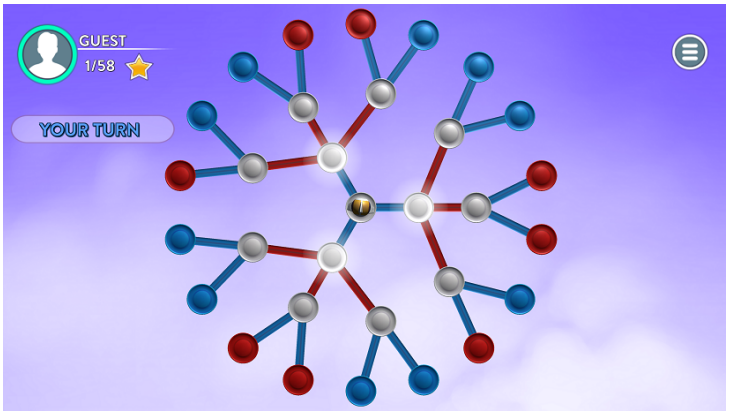


Figure 24: Non-tree 3.3.2

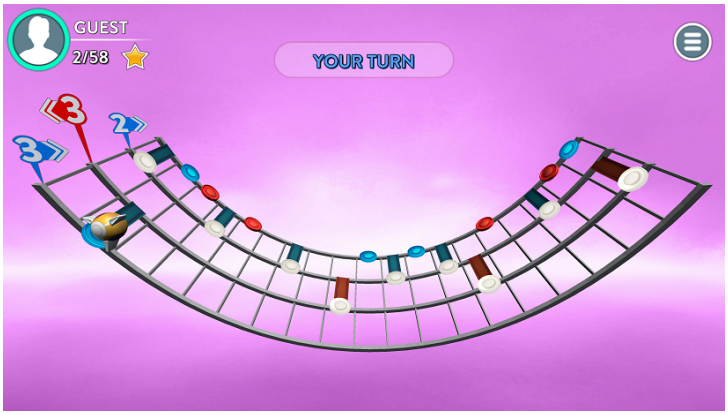


Figure 25: Tree 3.3.3

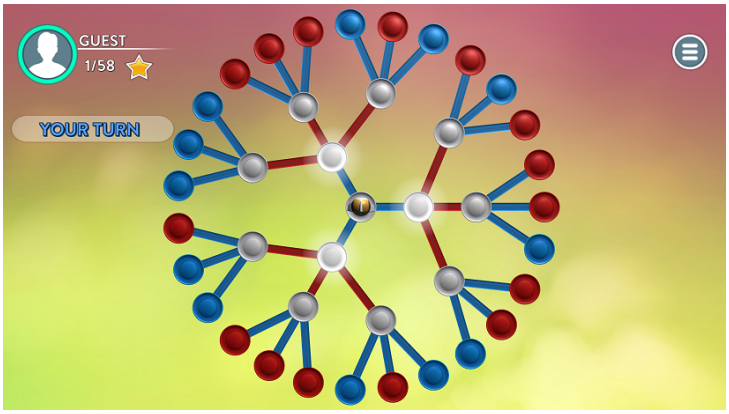


Figure 26: Non-tree 3.3.3

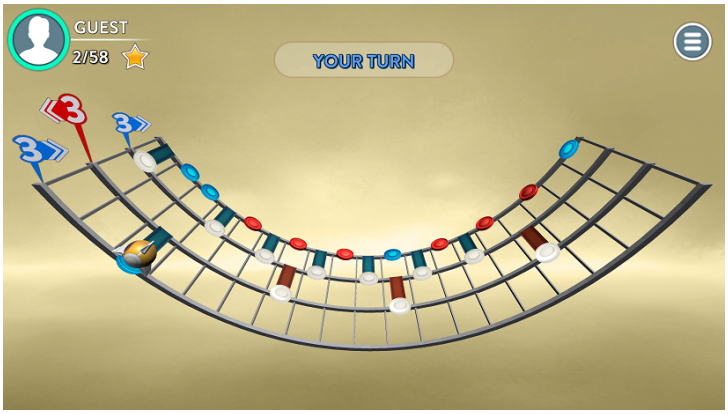


Figure 27: Tree 4.2.2

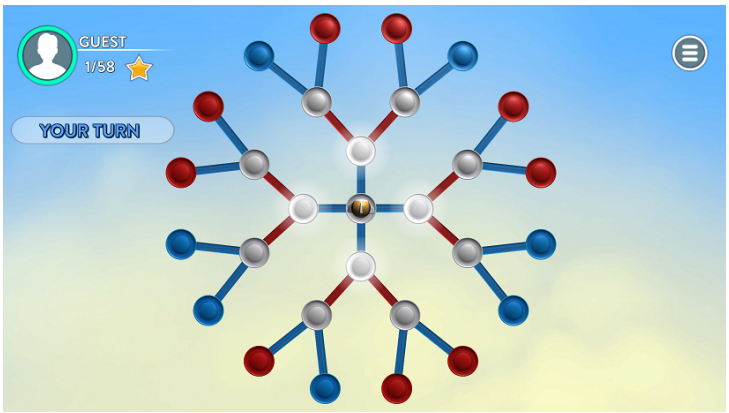


Figure 28: Non-tree 4.2.2

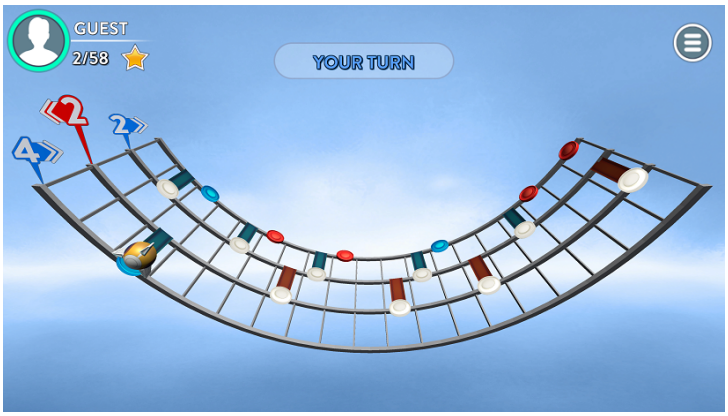


Figure 29: Tree 2.2.2.2

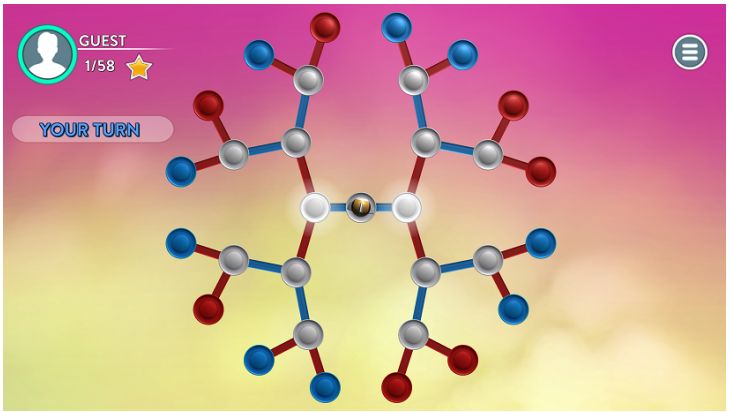


Figure 30: Non-tree 2.2.2.2

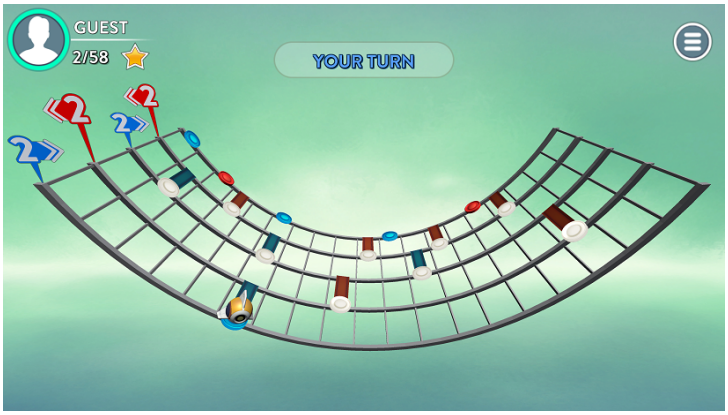


Figure 31: Tree 3.2.2.2

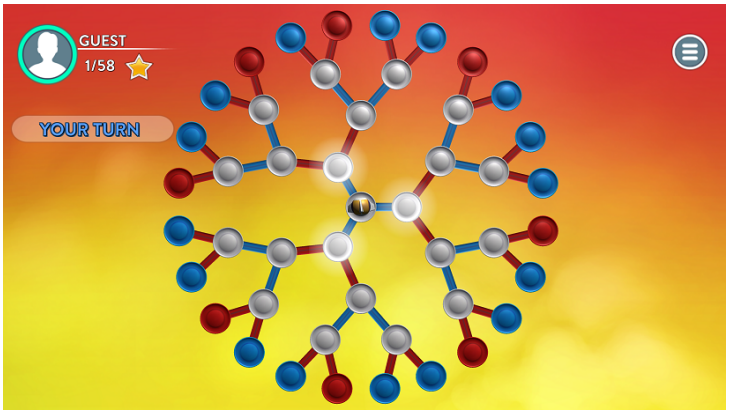


Figure 32: Non-tree 3.2.2.2

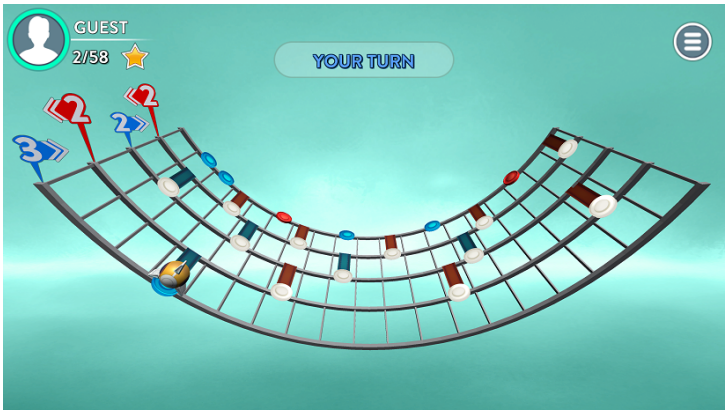


Figure 33: Tree 4.2.2.2

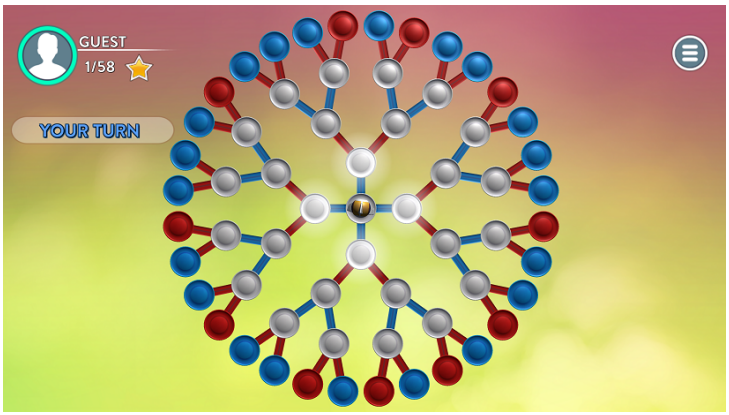


Figure 34: Non-tree 4.2.2.2

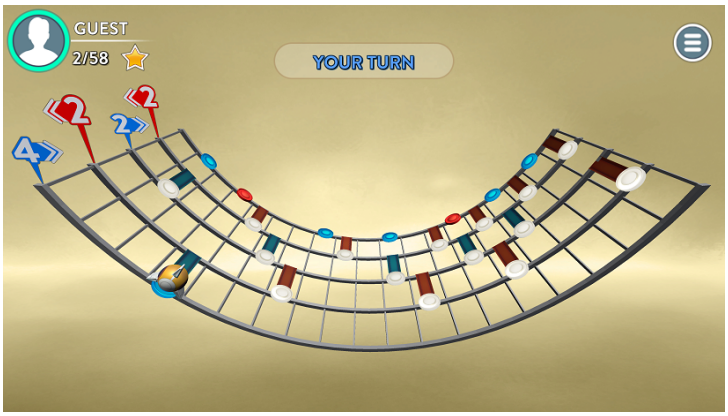


Figure 35: Tree 2.3.2.2

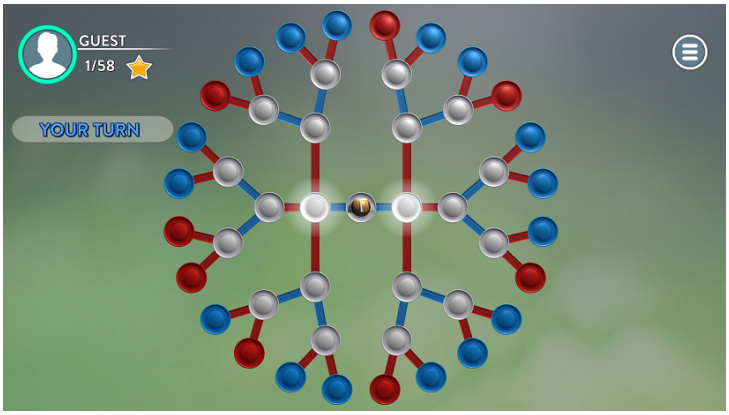


Figure 36: Non-tree 2.3.2.2

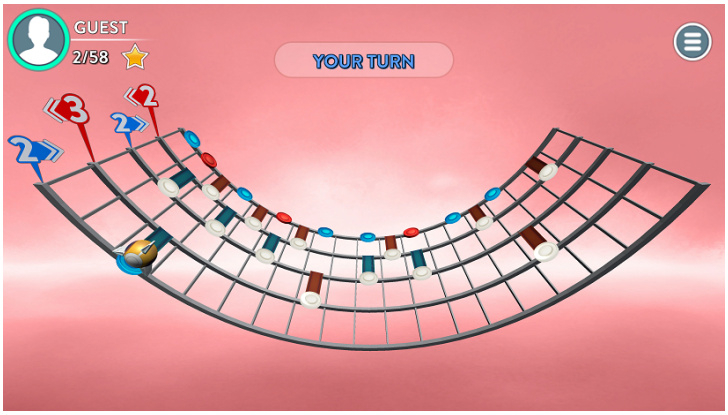


Figure 37: Tree 2.4.2.2

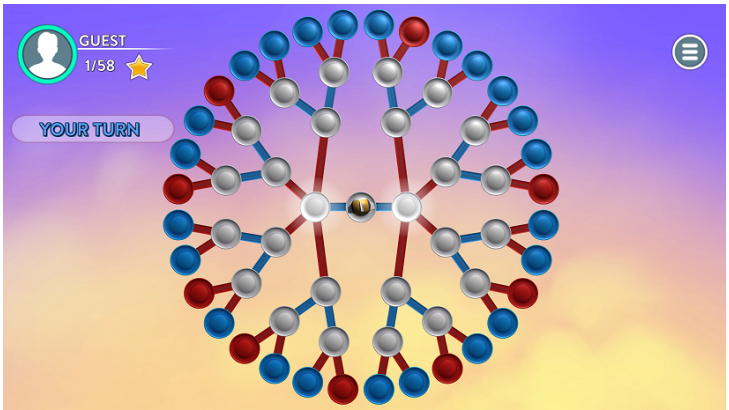


Figure 38: Non-tree 2.4.2.2

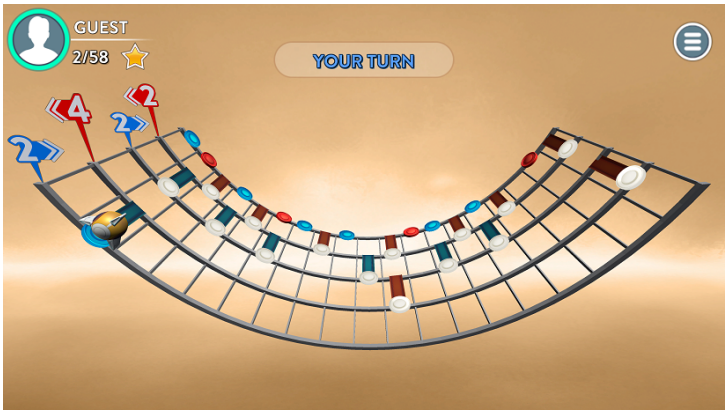


Figure 39: Tree 2.2.3.2

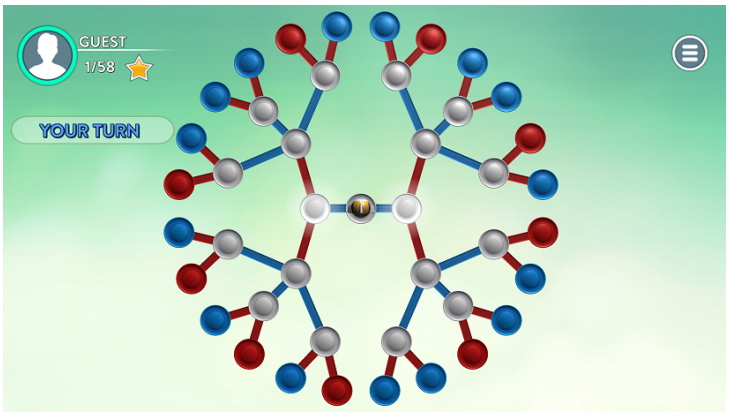


Figure 40: Non-tree 2.2.3.2

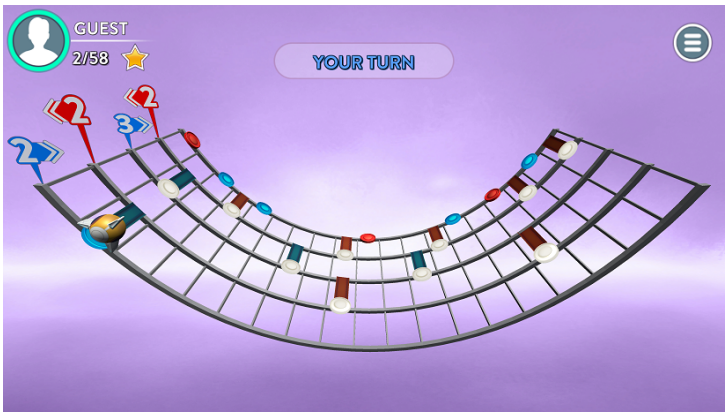


Figure 41: Tree 2.2.4.2

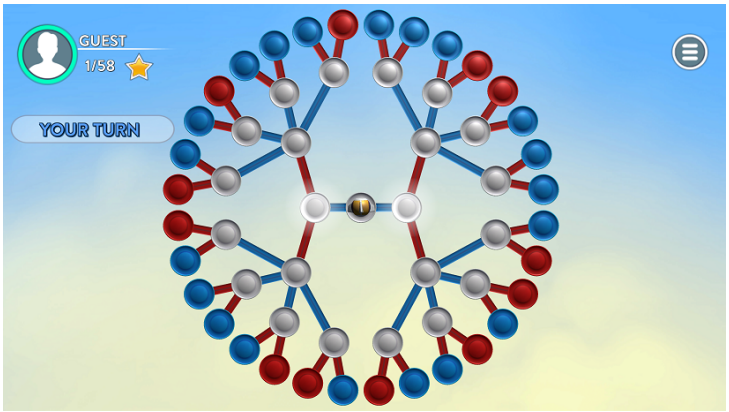


Figure 42: Non-tree 2.2.4.2

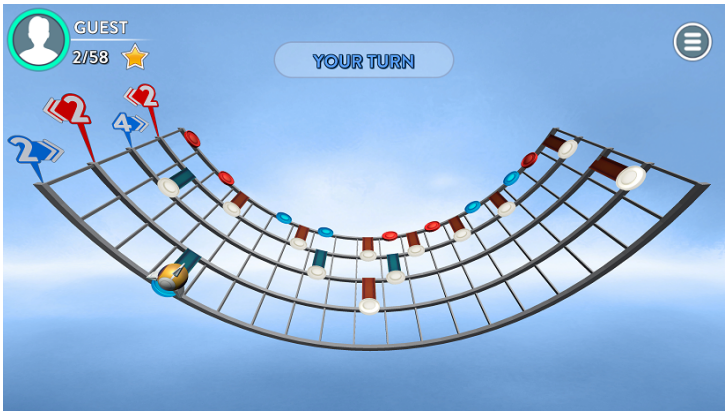


Figure 43: Tree 2.2.2.3

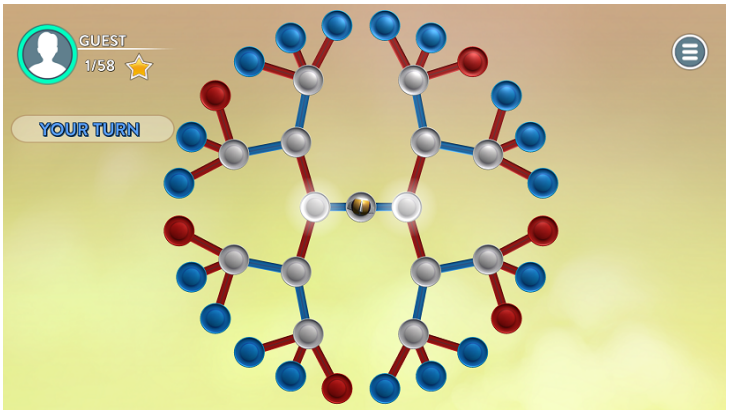


Figure 44: Non-tree 2.2.2.3

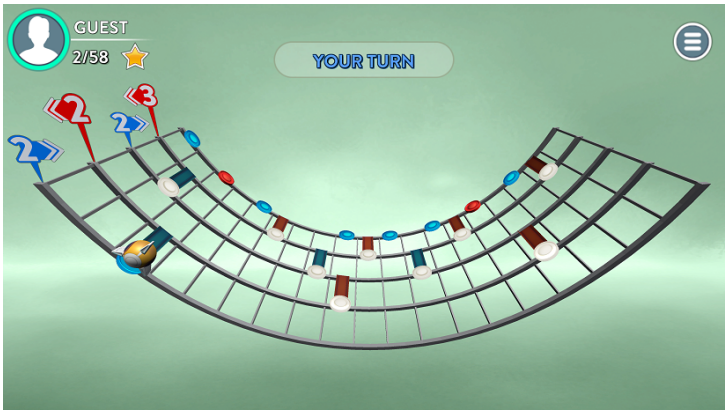


Figure 45: Tree 2.2.2.4

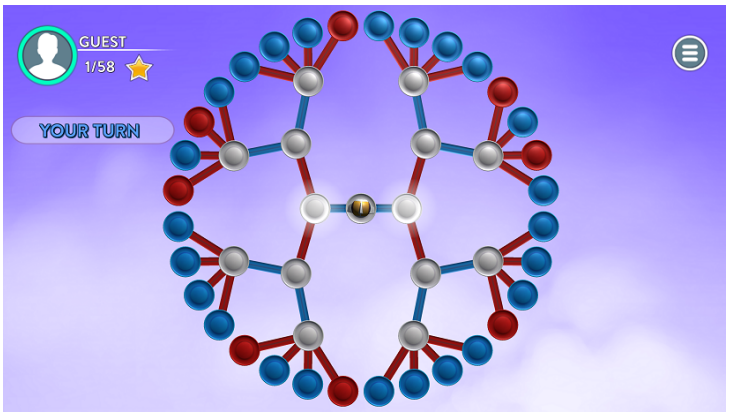


Figure 46: Non-tree 2.2.2.4

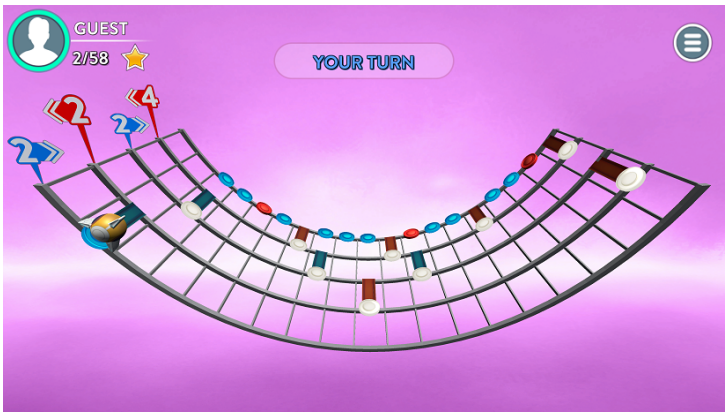


Figure 47: Tree 2.2.2.2.2

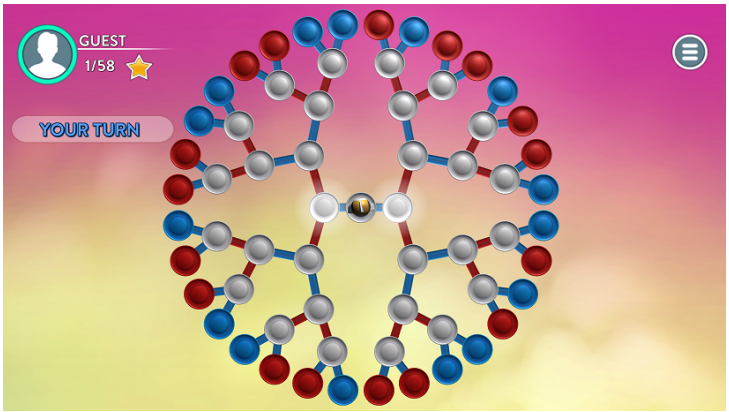


Figure 48: Non-tree 2.2.2.2.2

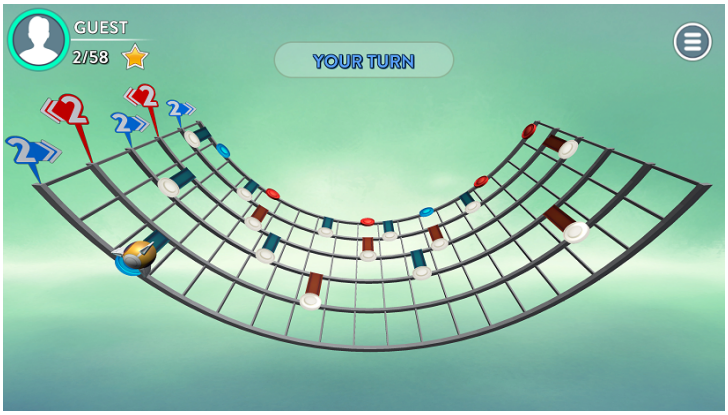


Figure 49: Tree 3.2.2.2.2

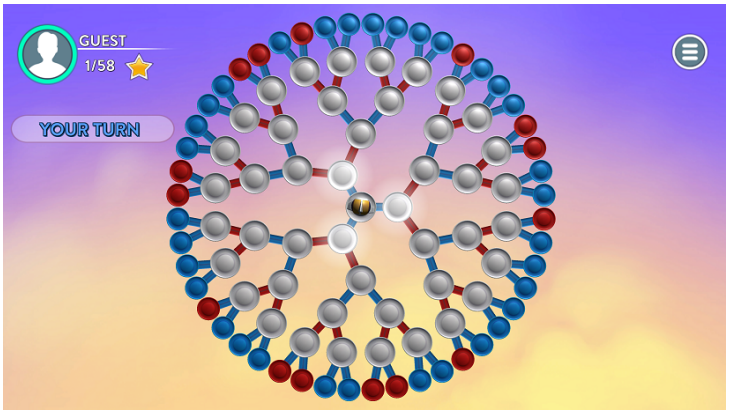


Figure 50: Non-tree 3.2.2.2.2

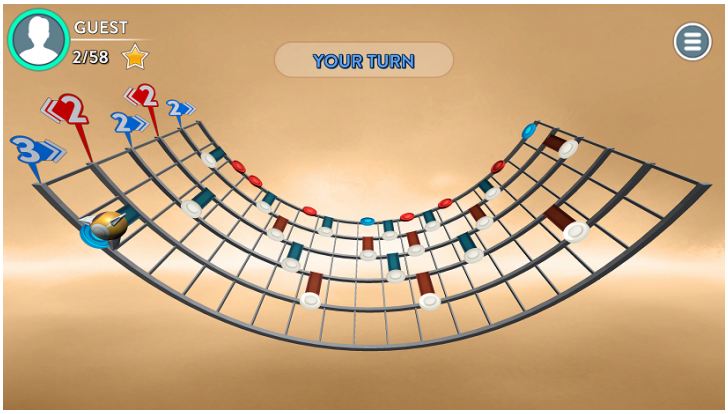


Figure 51: Tree 4.2.2.2.2

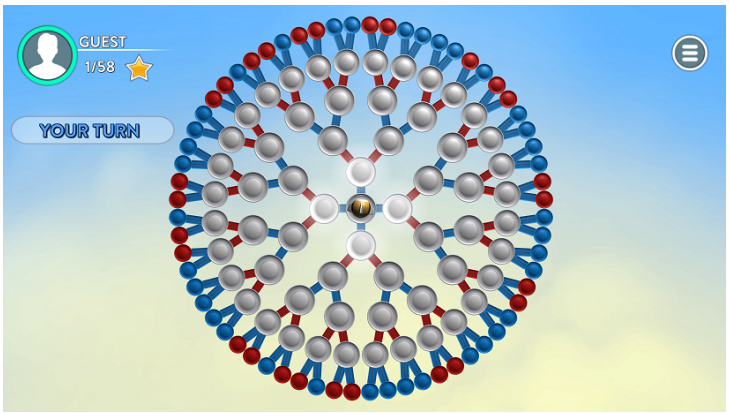


Figure 52: Non-tree 4.2.2.2.2

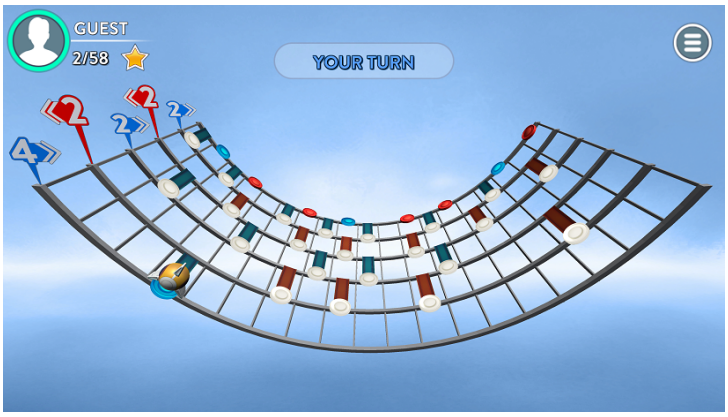


Figure 53: Tree 2.2.2.2.2.2

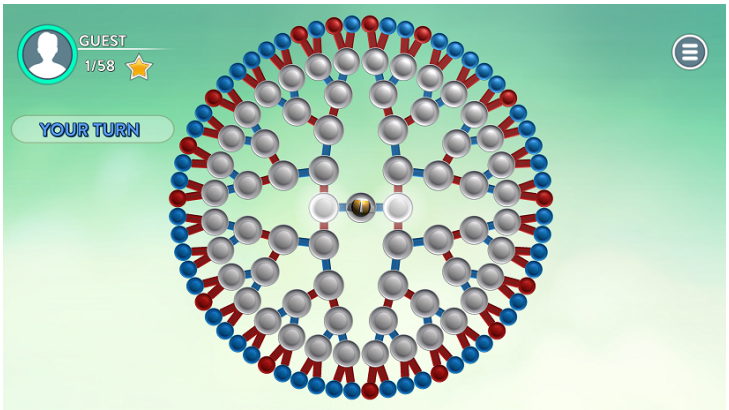
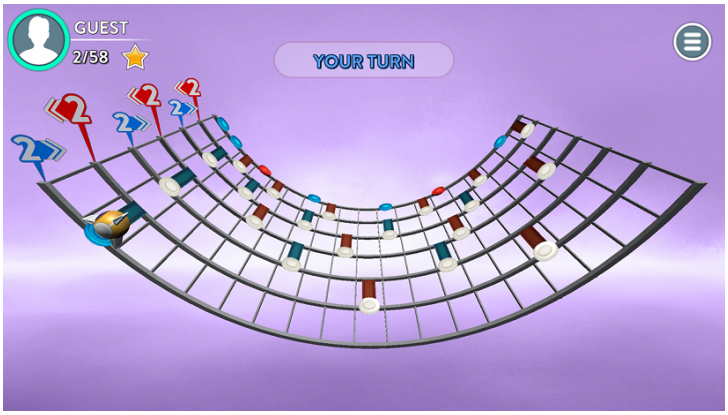


Figure 54: Non-tree 2.2.2.2.2.2



B Appendix B: *Blues and Reds* as a mobile app

In this appendix, we describe *Blues and Reds* as a mobile app; that is, we look at *Blues and Reds* from the gaming perspective. The appendix is especially useful for those who have not played *Blues and Reds*.

Blues and Reds is a puzzle mobile app consisting of 58 interactions divided into 10 chapters. Each interaction is a two-person, zero-sum finite dynamic game with perfect and complete information in which the subject plays against Artificial Intelligence (AI). There are nine regular chapters and one Immortal chapter; shortly, we explain the difference.

B.1 Main menu and chapter menu

Upon installing and opening *Blues and Reds* for the first time, the main menu screen with 10 chapters opens up (Figure 55). Only one chapter, called the Immortal chapter, is available, while all the remaining nine chapters are locked. Subjects enter a chapter by clicking on the chapter icon.

Figure 55: Main menu (initial view).

Once the subject completes the Immortal chapter, then the next chapter becomes available and the main menu changes. Figure 56 depicts the main menu after the subject completes the Immortal chapter and one regular chapter.

Figure 56: Main menu after completing the Immortal chapter and one regular chapter.

Each chapter consists of a specific number of interactions: there are four interactions in the Immortal chapter (see the Immortal chapter menu in Figure 57), and six interactions in every regular chapter (see the regular chapter menu Figure 58).

Figure 57: Initial menu of the Immortal chapter.

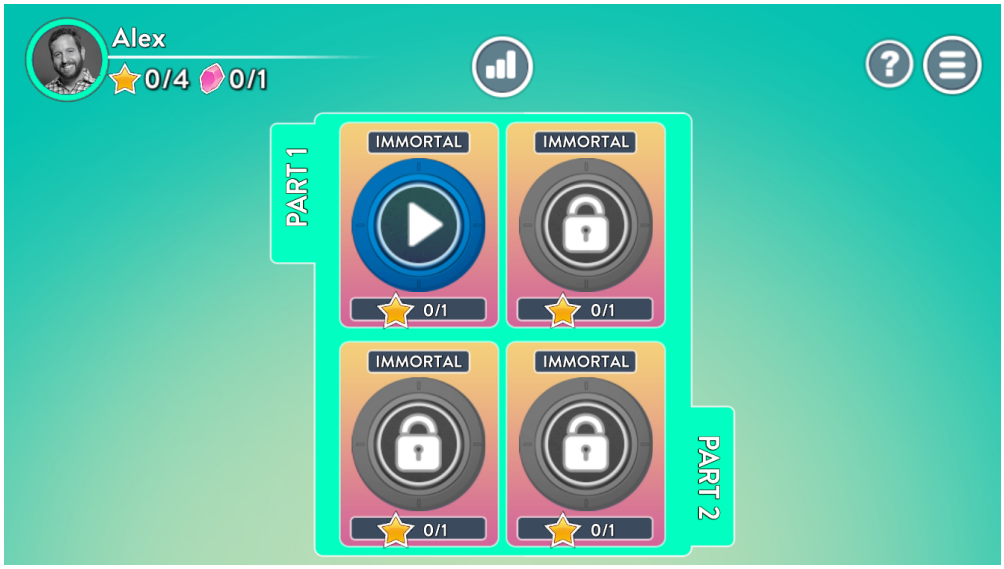
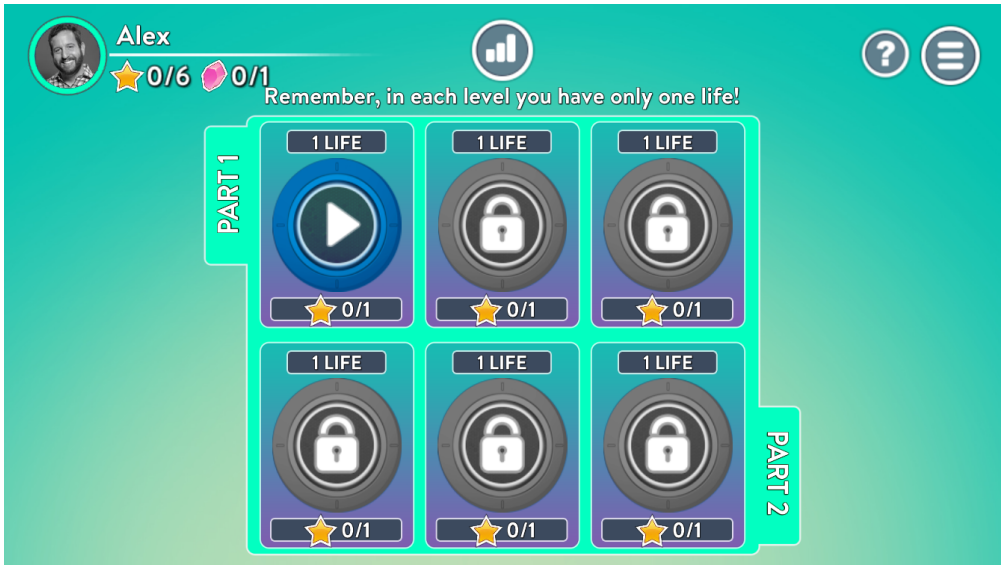


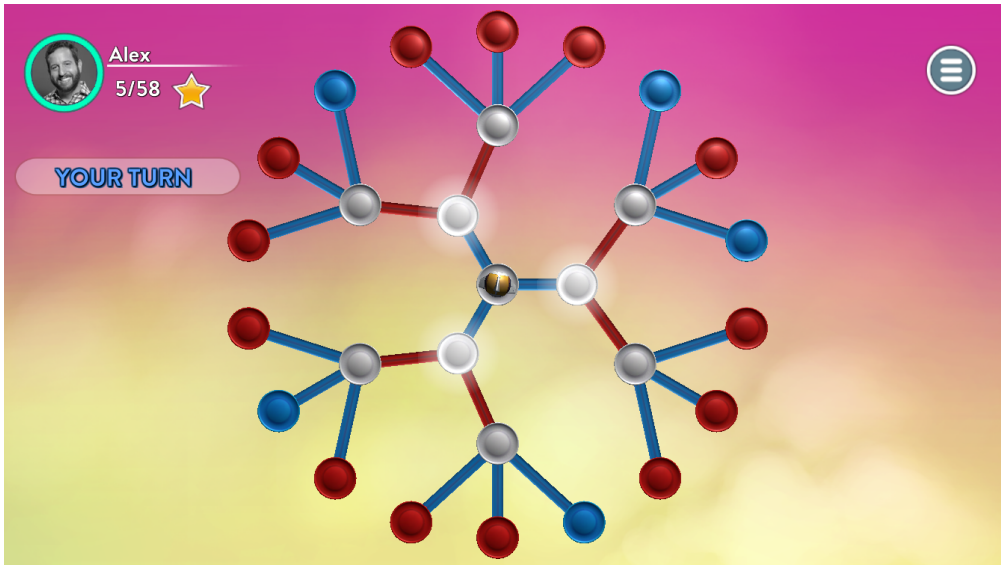
Figure 58: Initial menu of a regular chapter.



B.2 Tree interactions and non-tree interactions

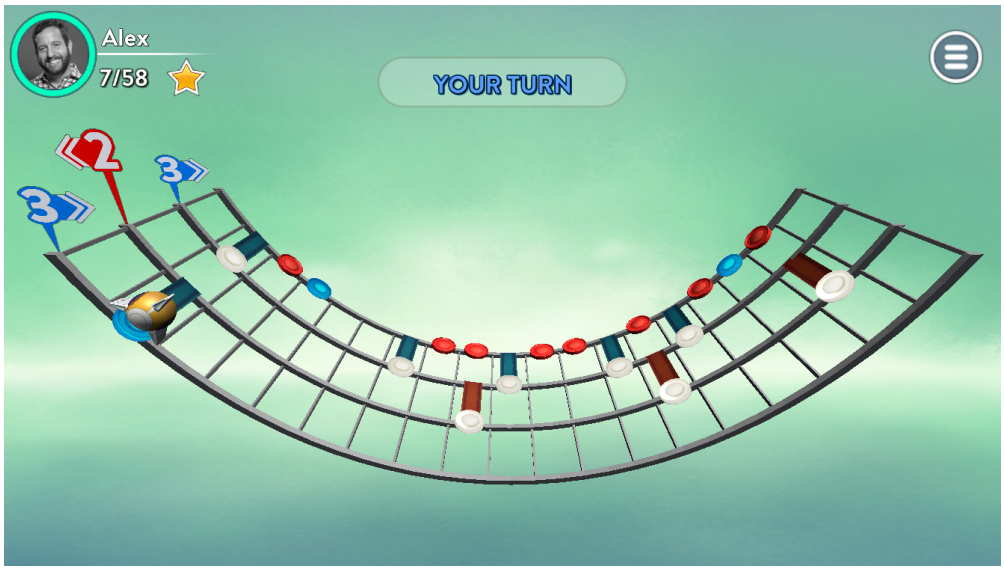
Each chapter menu – Immortal or regular – consists of two rows. In the top row, there are *tree* interactions – these are the interactions that have a structure of a game-theoretic tree. An example of a tree interaction is depicted in Figure 59.

Figure 59: Tree interaction.



In the bottom row, there are *non-tree* interactions – while they, as tree interactions, are also finite dynamic games with complete and perfect information, they have a different, more convoluted structure. An example of a non-tree interactions is depicted in Figure 60.

Figure 60: Non-tree interaction.



Note that the tree interaction in Figure 59 and non-tree interaction in Figure 60 are 3.2.3 interactions.

Each interaction – whether tree or non-tree – has the same basic rules and objectives. The subject and AI move the golden spherical object, called the RoboToken, across the blue (subject) and red (AI) bridges. Choices are made in turns with the subject moving first.

When it is the subject's turn, all of her available choices are highlighted. Each time a subject makes a choice, she has to confirm her move; this feature prevents the subject from making choices she did not intend to do.

The interaction ends when the RoboToken lands on a blue node (subject wins) or a red node (subject loses). If the subject wins, then she is awarded a star. If she loses, then she does not gain a star, but there is no additional punishment. If the subject wins all interactions in a given chapter, she also gains a diamond.

In a tree interaction, the subject and AI choose in turns which bridge the RoboToken is to cross. The subject does it by selecting a node that she wants the RoboToken to fly onto. Then, AI moves the RoboToken across one of the red bridges from the node that the subject selected in the previous round. And so on. Every tree interaction starts with the RoboToken

initially located at the center of the tree as in Figure 59.

A non-tree interaction consist of k rails; for example, $k = 4$ in Figure 60. Except for the last rail, at each rail there is a blue or red arrow. This arrow indicates the maximum number of available moves and the direction of the move. The subject moves at rails with a blue flag, and AI moves at rails with a red flag.

For example, in Figure 60, in the first round, the subject can choose one of three nodes on the next rail to the right of the current position of the RoboToken. The subject chooses a node and confirms her choice. Then, the blue bridge drifts to that node and the RoboToken flies to the selected node.

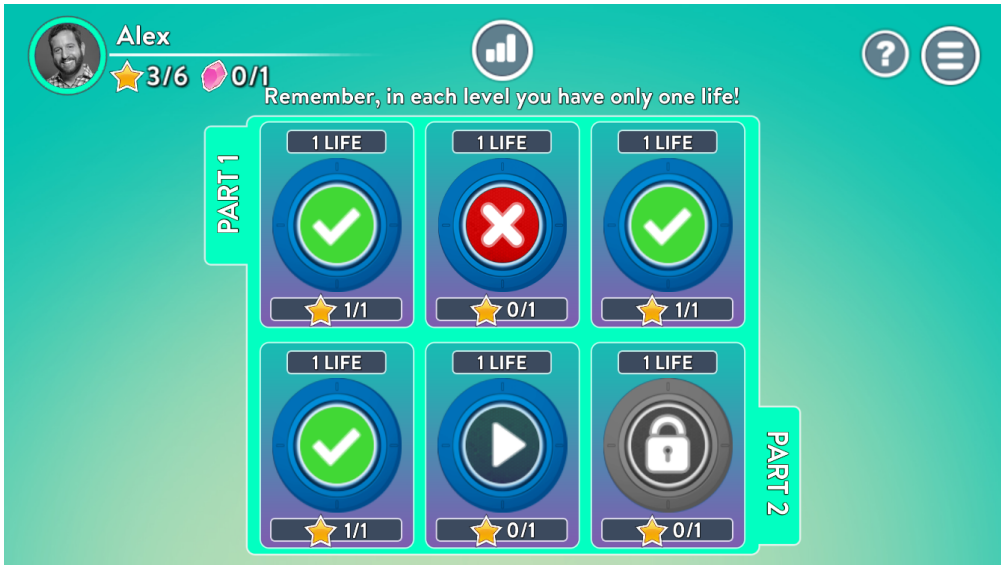
At the next round, AI selects one of two available nodes to the left of the node that RoboToken is at after the first round (i.e., the node that the subject selected in the first round). Finally, in the third round, it is again the subject who chooses a node that is among three nodes to the right of the node that RoboToken is at after two rounds (i.e., the node that AI selected in the second round).

B.3 Regular chapters

In every interaction of each regular chapter, there is only "one life per interaction," i.e., the subject can attempt each interaction only once. In addition, playing interactions in regular chapters, the subject is left on her own – no help or hint is provided.

In each regular chapter, the subject is unable to select which interaction she plays; she unlocks interaction by interaction. The next interaction becomes available only if the subject completes (wins or loses) the currently available interaction. In Figure 61, we depict the chapter menu after the subject has played all tree interactions (two wins, one loss) and one non-tree interaction (win); the second interaction in the row of non-tree interaction is the next available interaction.

Figure 61: Menu of a regular chapter after completing four interactions.



B.4 The Immortal chapter and text tutorial as experimental instructions

In the Immortal chapter, which consists of four interactions (Figure 57), the subject has infinitely many lives – that is, making mistakes has no negative consequences – and can play interactions in this chapter as many times as she wants. As already explained, this is not the case in the remaining nine regular chapters.

The Immortal chapter is mandatory and serves the role of hands-on experimental instructions. The subject must complete all the interactions in the Immortal chapter in order to progress in *Blues and Reds*. The goal is to force the subject to become familiar with the rules and objectives of *Blues and Reds*. The Immortal chapter is accessible at no cost and at any moment of playing *Blues and Reds*.

The Immortal chapter instructs the subject that she plays in two-person interactions against AI. The subject is told what must happen for her to win/lose. She also is instructed that her winning means that AI loses and vice versa (zero-sum payoffs without ties). Finally, the subject learns how to make choices while playing tree and non-tree interactions. Details of experimental instructions are provided below.

The Immortal chapter starts with only one interaction available to the subject as depicted in Figure 57. This first interaction is a tree interaction in which the subject does not really

make any choices because all the moves are indicated by the tutorial. The goal is to force the subject to learn what a tree interaction is and how to play it. When the subject wins the first interaction, she gains her first star and the second interaction becomes available.

In the second interaction, which is also a tree interaction, the subject solves a more complex tree. While she is not guided and can choose what she wants, the subject is provided with the hint option (accessible at no cost). In order to progress to the next interaction, it is necessary for the subject to win this interaction and collect her second star; if the subject loses her second tutorial interaction, then she is forced to repeat it until she wins.

Figures 62–69 depict the chronological order of screenshots in the two tutorial tree interactions.

Figure 62: First tutorial tree (screenshot 1).

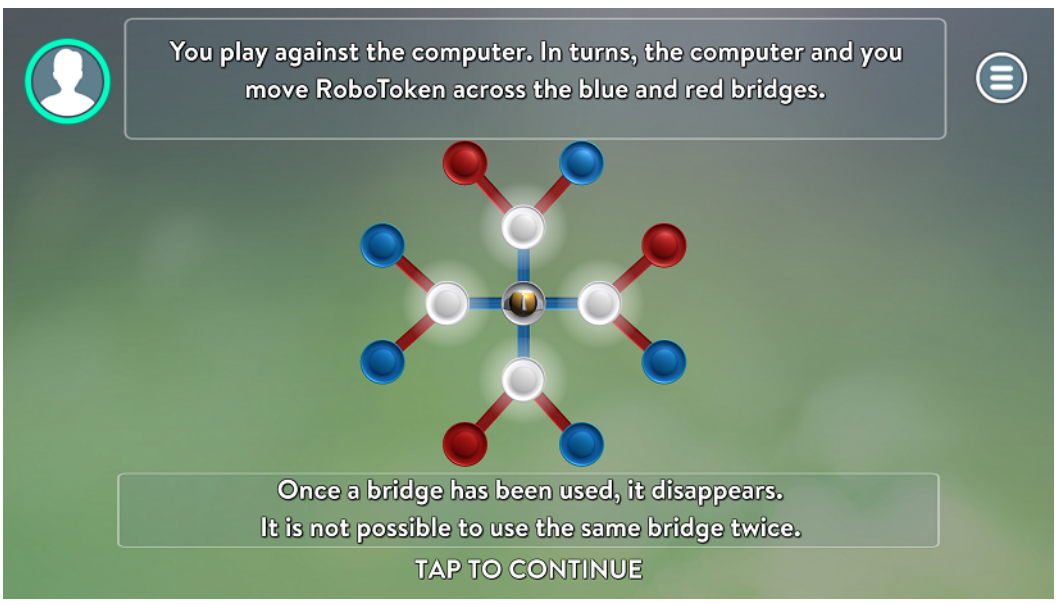


Figure 63: First tutorial tree (screenshot 2).

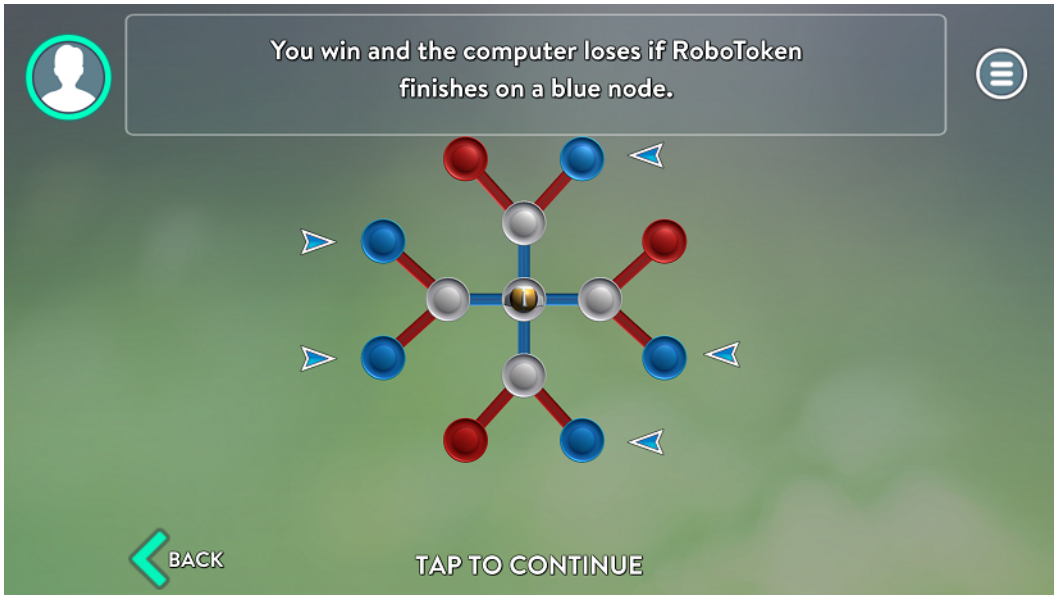


Figure 64: First tutorial tree (screenshot 3).

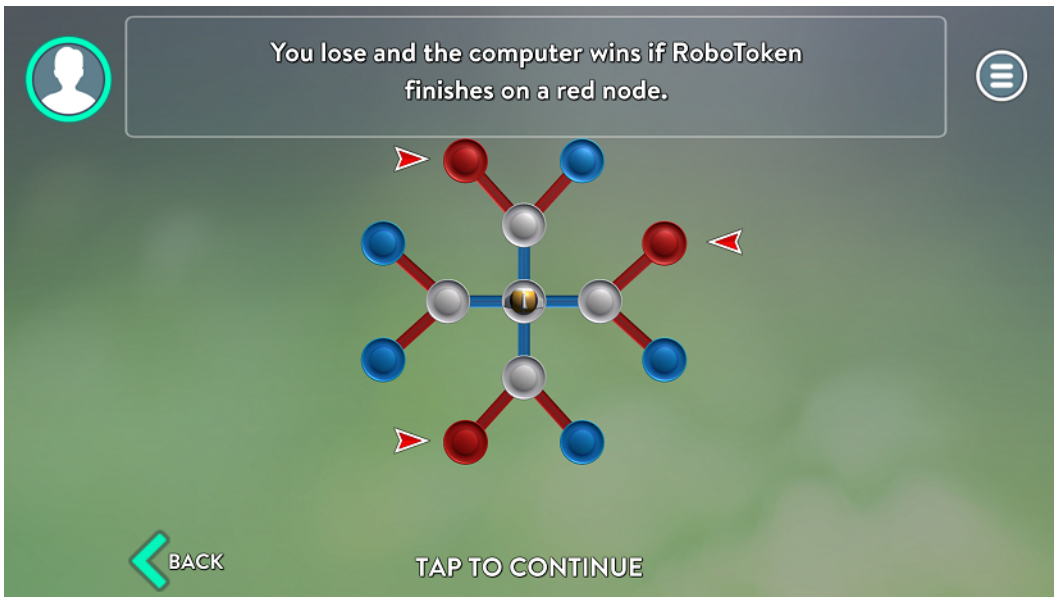


Figure 65: First tutorial tree (screenshot 4).

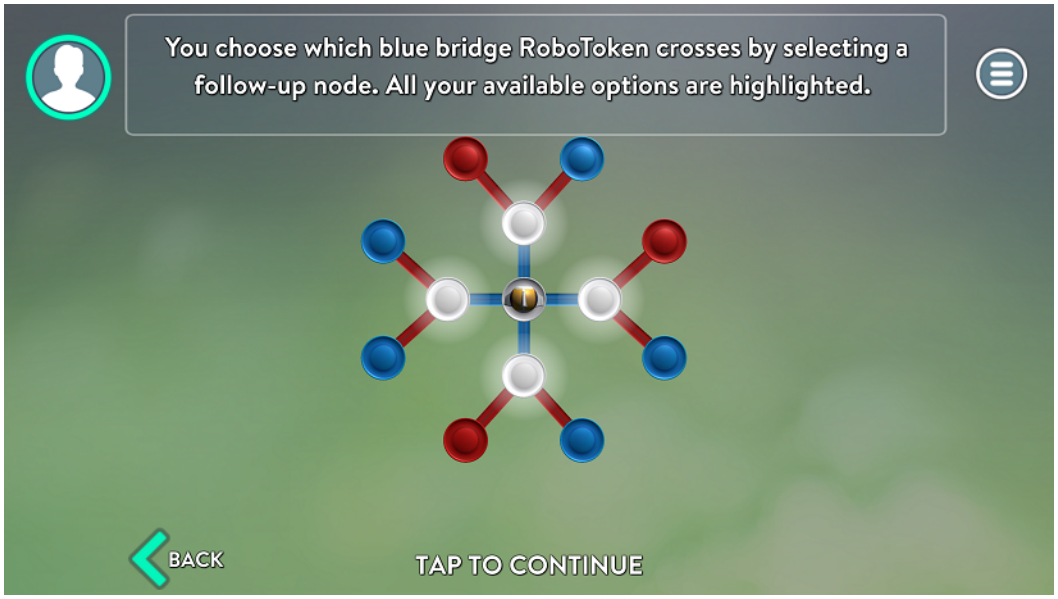


Figure 66: First tutorial tree (screenshot 5).

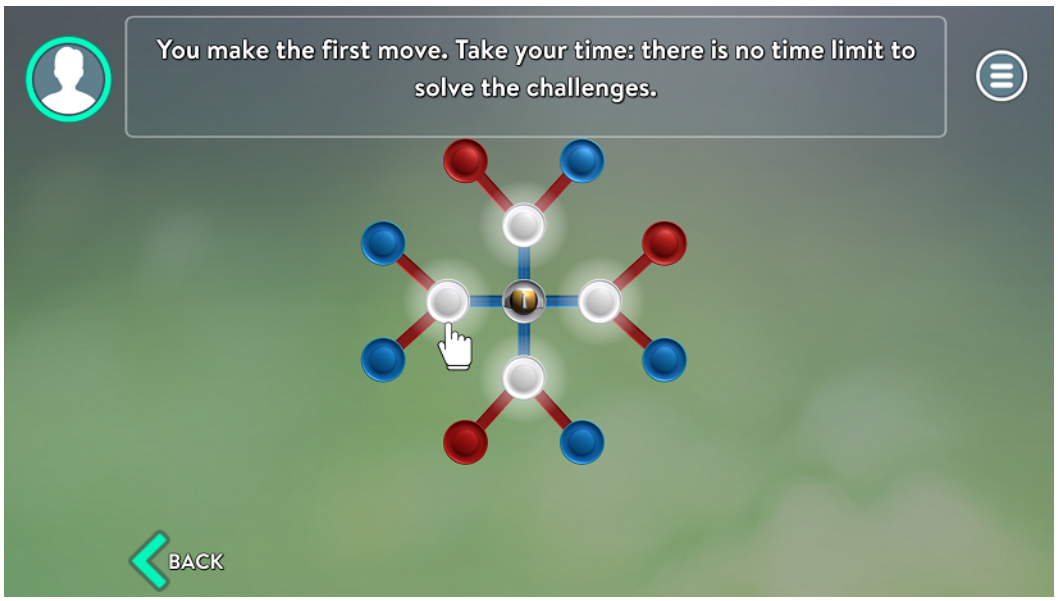


Figure 67: First tutorial tree (screenshot 6).

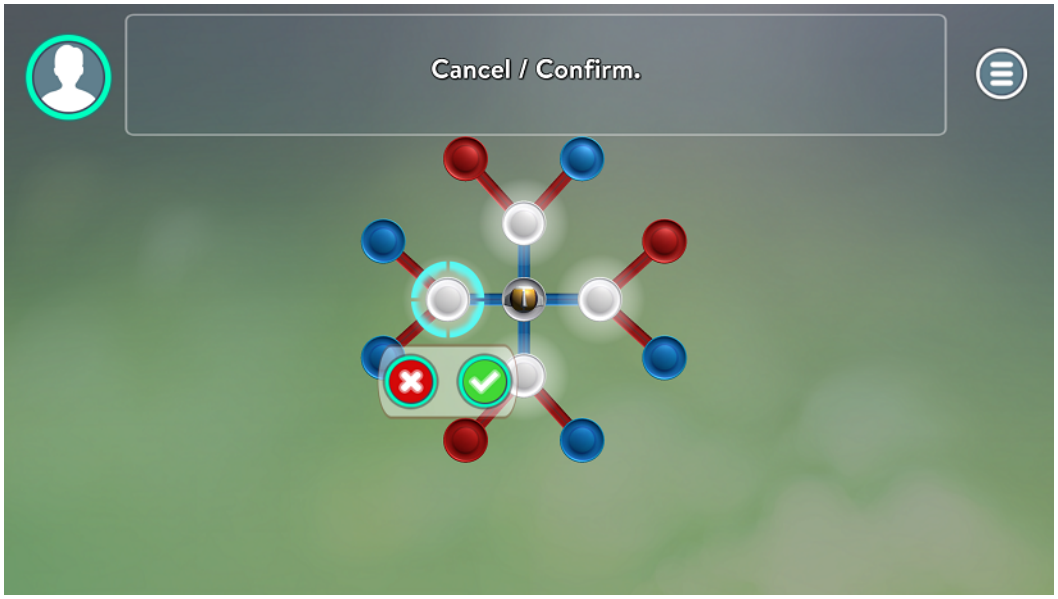


Figure 68: First tutorial tree (screenshot 7).

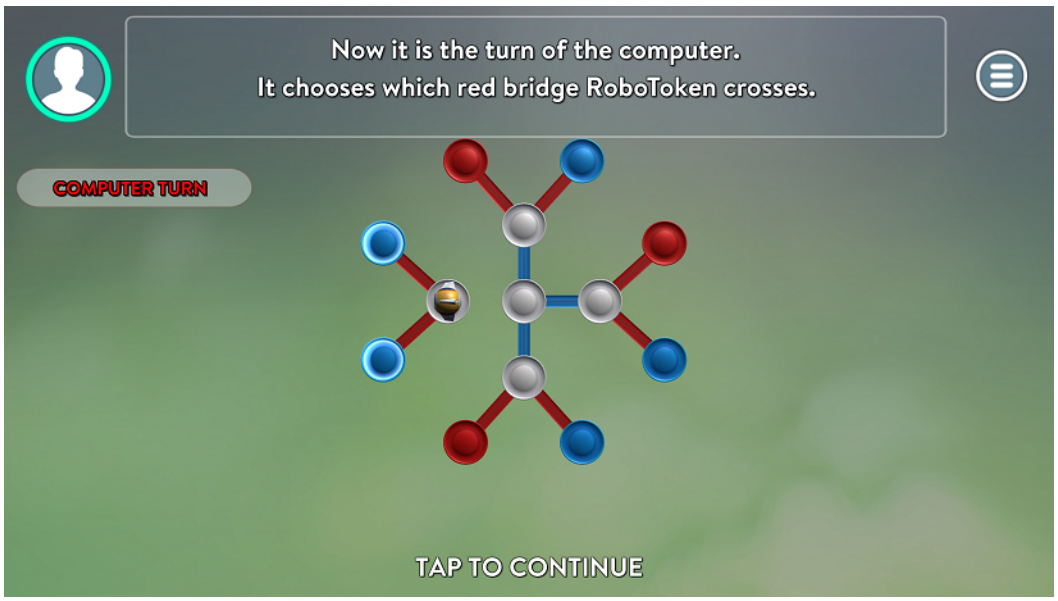
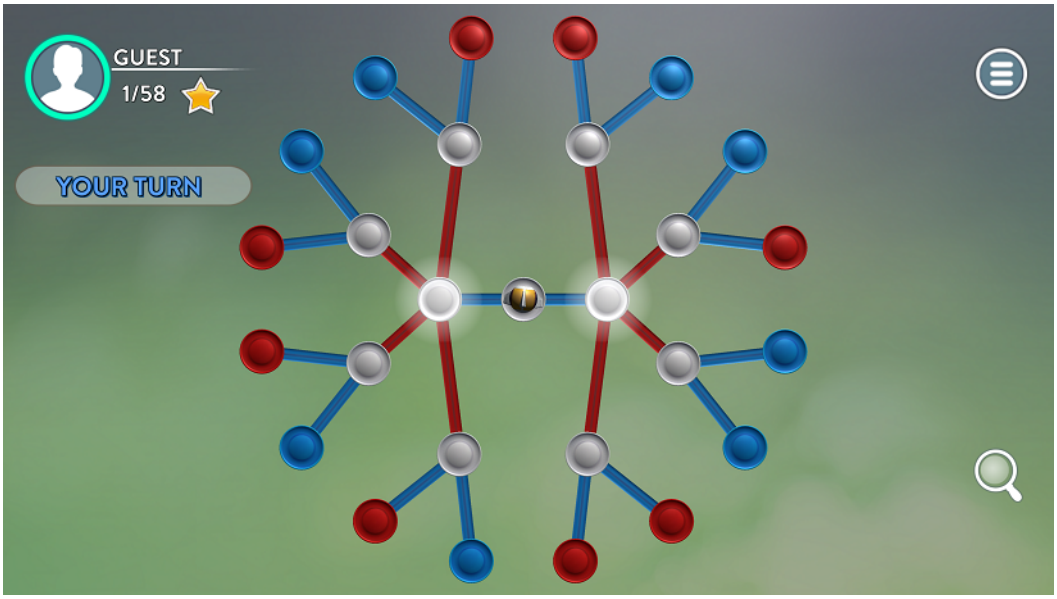


Figure 69: Second tutorial tree.



When the subject wins the two tree tutorial interactions, she then plays two tutorial non-tree interactions. The first tutorial non-tree interaction, as the first tutorial tree interaction, forces choices upon the subject to explain a non-tree interaction. The subject wins this interaction and collects her third star.

In the second tutorial non-tree interaction, as with the second tutorial tree interaction, the subject makes her own choices but is provided with the costless hint option. The subject must win this interaction in order to progress; if she loses, she repeats the interaction as many times as it is necessary for her to win.

Figures 70–79 depict the chronological order of screenshots in the two tutorial non-tree interactions.

Figure 70: First tutorial non-tree (screenshot 1).

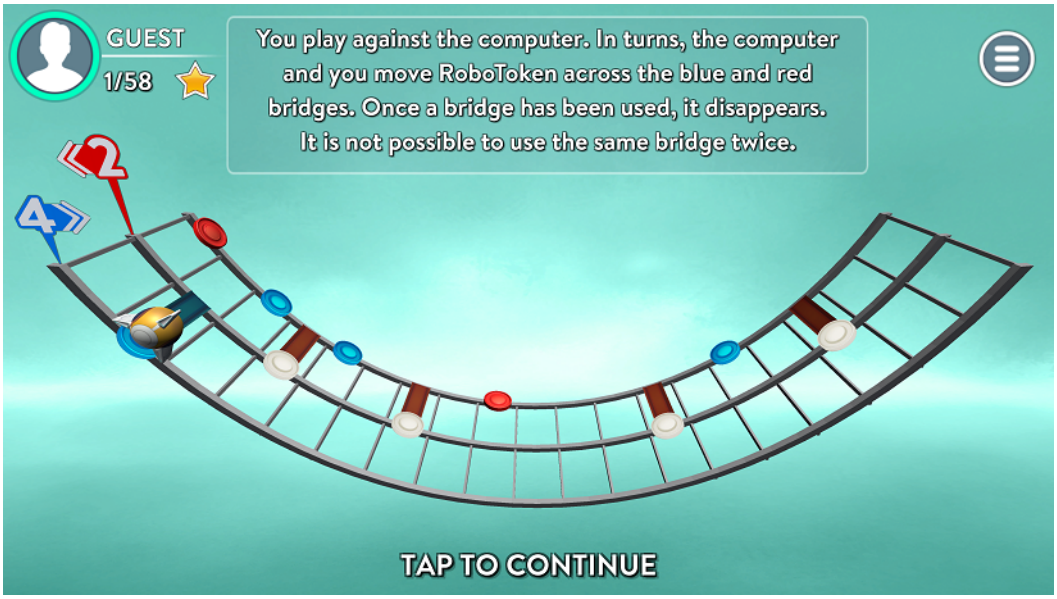


Figure 71: First tutorial non-tree (screenshot 2).

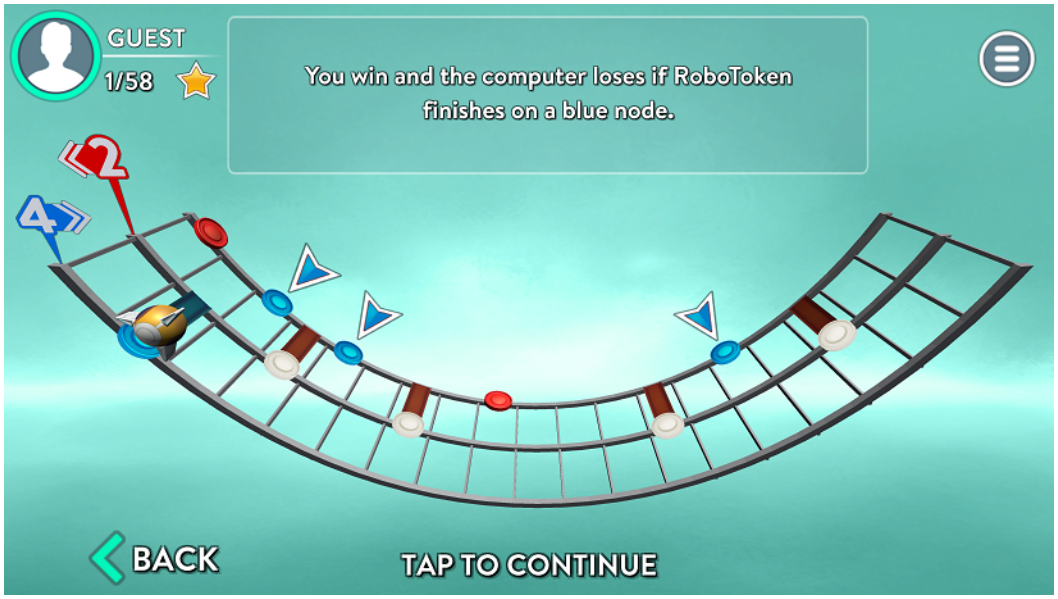


Figure 72: First tutorial non-tree (screenshot 3).

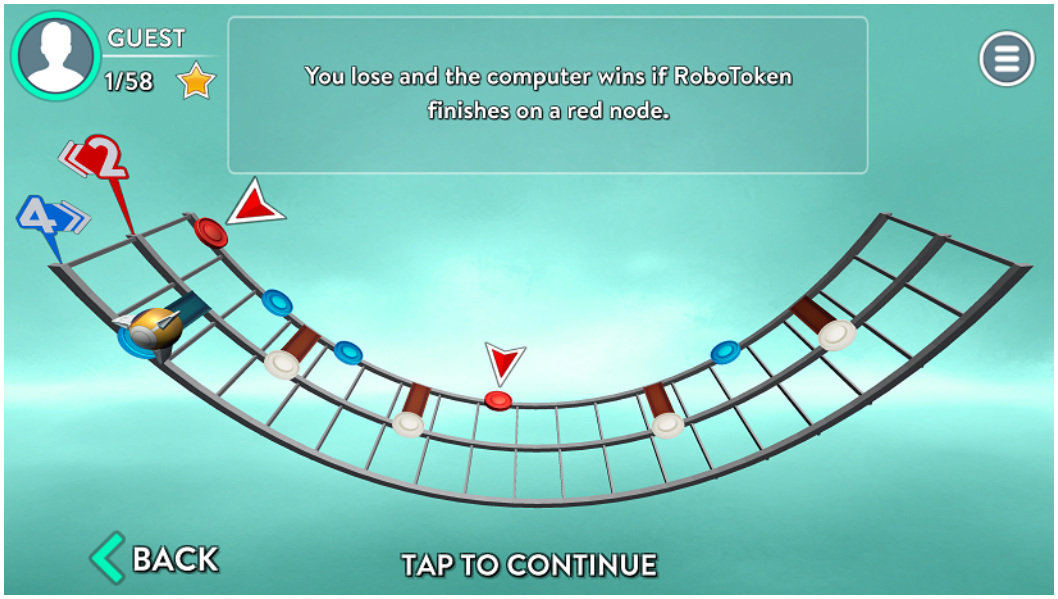


Figure 73: First tutorial non-tree (screenshot 4).

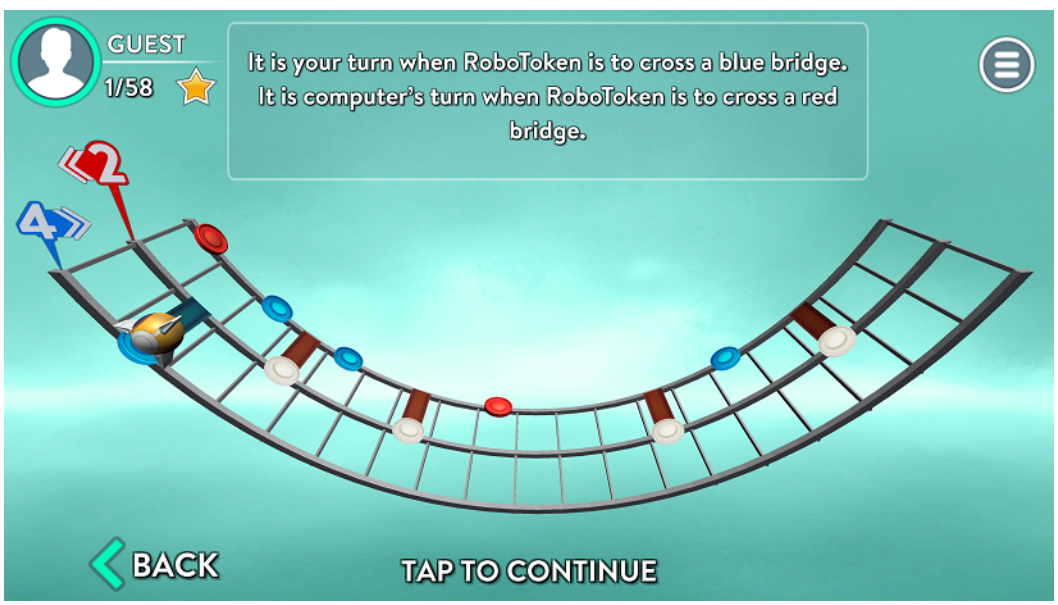


Figure 74: First tutorial non-tree (screenshot 5).

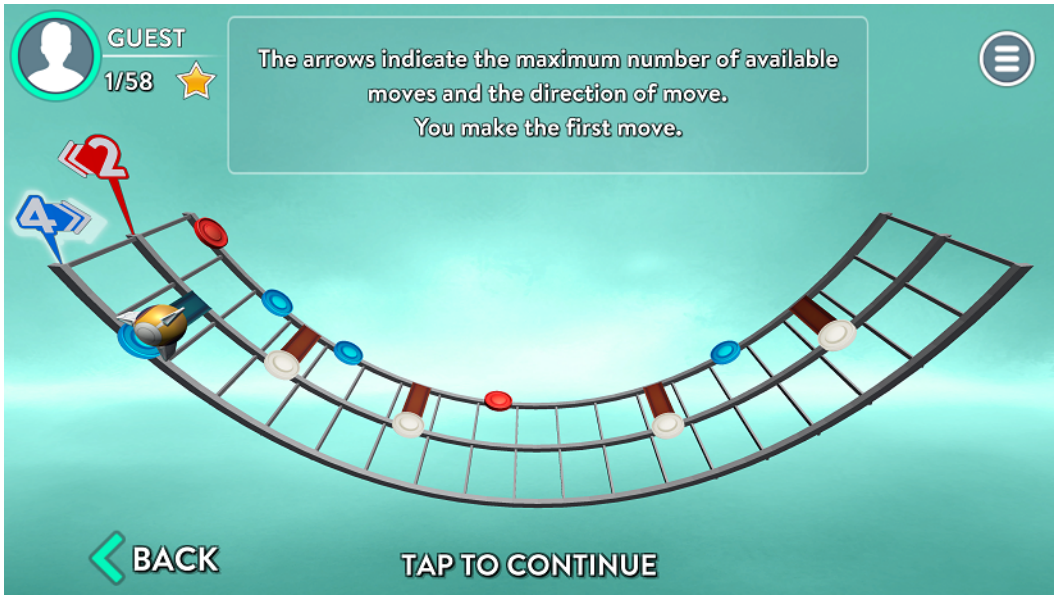


Figure 75: First tutorial non-tree (screenshot 6).

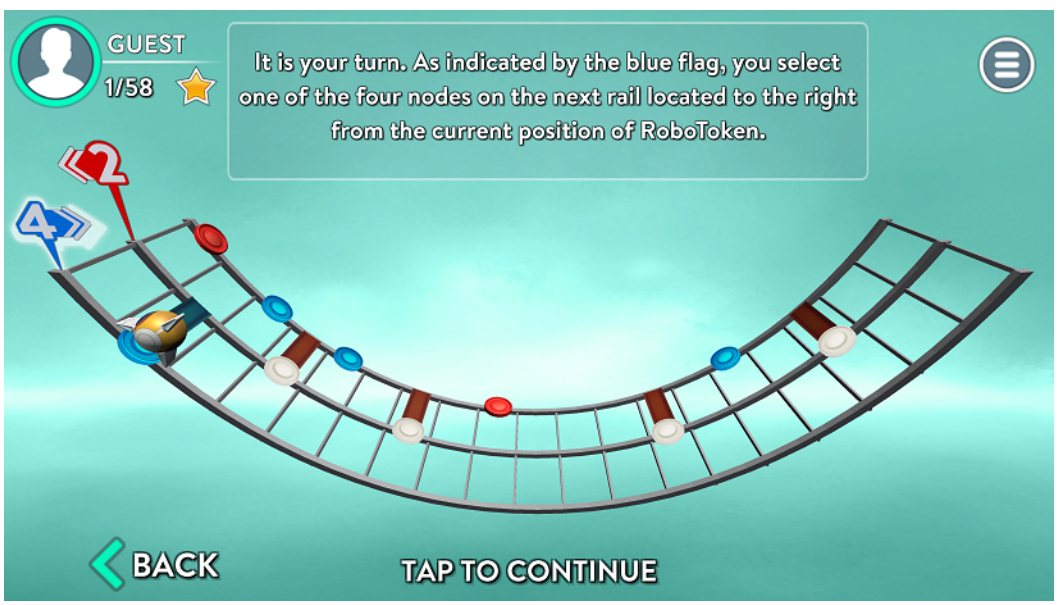


Figure 76: First tutorial non-tree (screenshot 7).

Figure 77: First tutorial non-tree (screenshot 8).

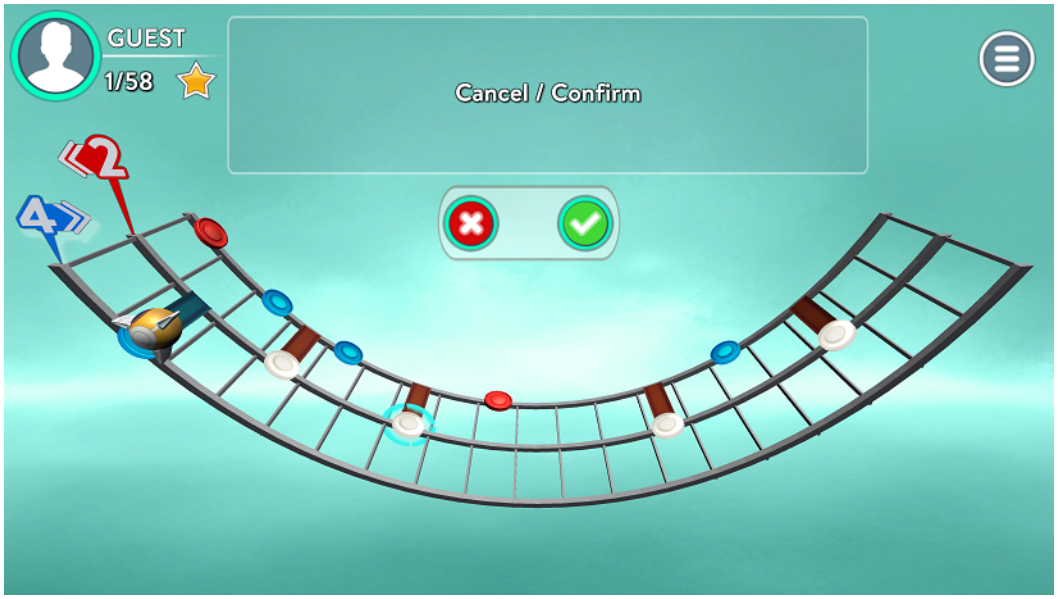


Figure 78: First tutorial non-tree (screenshot 9).

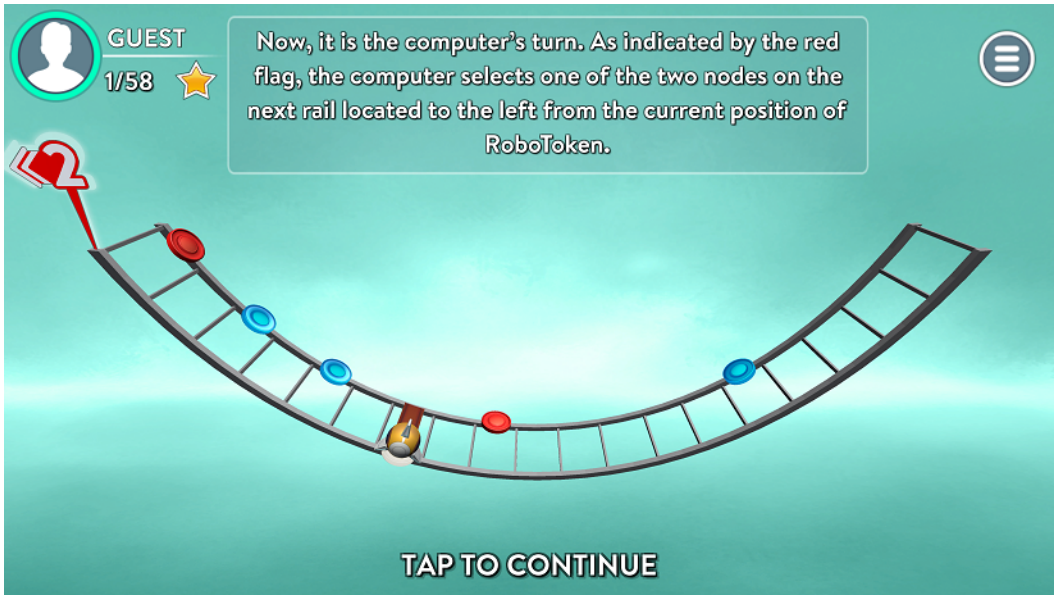
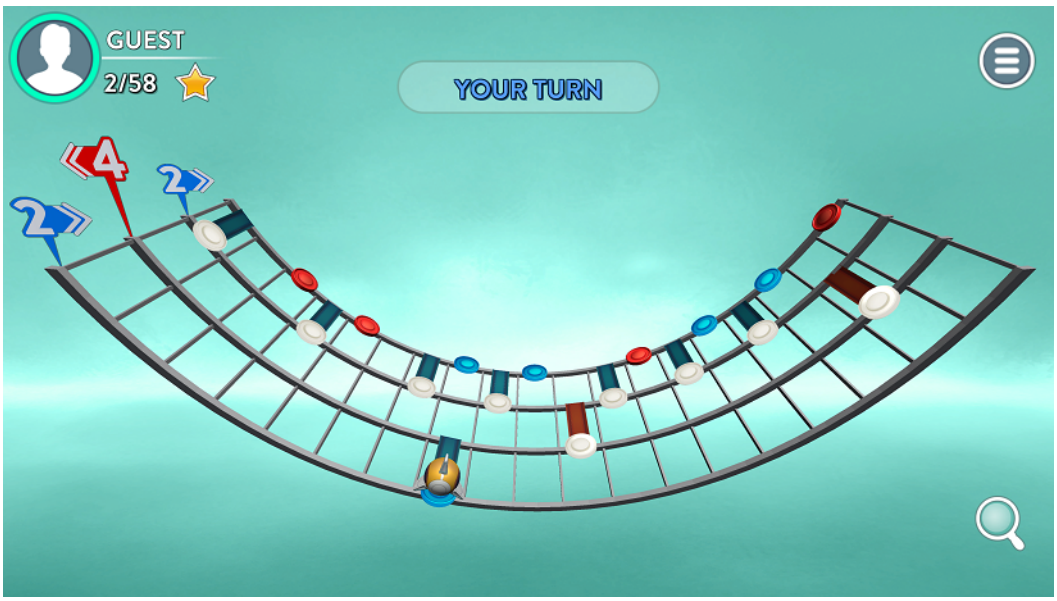


Figure 79: Second tutorial non-tree.



The Immortal chapter ends with the subject winning all four interactions and collecting four stars and one diamond. When the subject completes the mandatory tutorial and wants to play the interactions in the Immortal chapter again, she can then choose whatever interaction she wants as all interactions in the Immortal chapter remain unlocked.

While playing *Blues and Reds*, subjects can also access the text tutorial which is always available upon clicking the question mark button. The text tutorial is not mandatory for the subjects to read. Below, we provide the complete text tutorial. (Note that the term "game" in the text tutorial refers to *Blues and Reds* as a mobile game rather than to a game in a game-theoretic sense.)

Game

The game consists of ten chapters. Each chapter consists of Part 1 and Part 2. In the Immortal Chapter, each part consists of two levels. In all other chapters, each part consists of three levels.

RoboToken

In turns, you and the computer decide where to move RoboToken.

How to win

You win if RoboToken finishes on a blue node. You lose if RoboToken finishes on a red node. If you win a level, you collect a star. If you win all levels in a chapter, you collect a diamond.

Playing levels in Part 1

Each level in Part 1 is a turn-based game against the computer. You start by selecting which blue bridge RoboToken crosses. All your available moves are highlighted. Once you make a choice, the computer chooses which red bridge to cross. This process continues until RoboToken ends up in a blue or red node.

Playing levels in Part 2

Each level in Part 2 is a turn-based game against the computer. You start by moving a blue bridge for RoboToken to cross. The blue arrows indicates the maximum number of available moves and the direction of the move. All your available moves are highlighted. Once you make a choice, it is the computer that moves a red bridge for RoboToken to cross. The red arrow indicates the maximum number of available moves and the direction of the move for the computer. This process continues until RoboToken ends up in a blue or red node.

Immortal Chapter

The first chapter is the Immortal Chapter. You can play the levels in this chapter as many times as you want. If you make a mistake, you can play again. When you play other chapters, you can always go to the main menu to play the Immortal Chapter again and then return to your level.

One Chapter = One Life

Except for the Immortal Chapter, in each chapter you have ONLY one life per level. Take

your time, there is no time limit to complete the levels. To avoid mistakes, you will need to confirm each move. Once you confirm your move, you will be unable to re-do it.

Tutorial

This tutorial is available at any moment of the game. If you open the tutorial, then the progress of your game will be saved. After reading the tutorial, you return to where you were in the game.

Progress

When you play, your progress is automatically saved. If you open the tutorial, return to the main menu, or leave the game, you will be returned to where you left of. Additionally, if you delete and re-download the game, you will be returned to where you left off.

Results

In this game, you compare your skills with other people. If you provide your information (country, age, and gender), then you can compare your results against your peer groups. Your results will be available only if you complete (win or lose) all levels.

To learn more: *www.bluesandreds.com*