



Ca' Foscari
University
of Venice

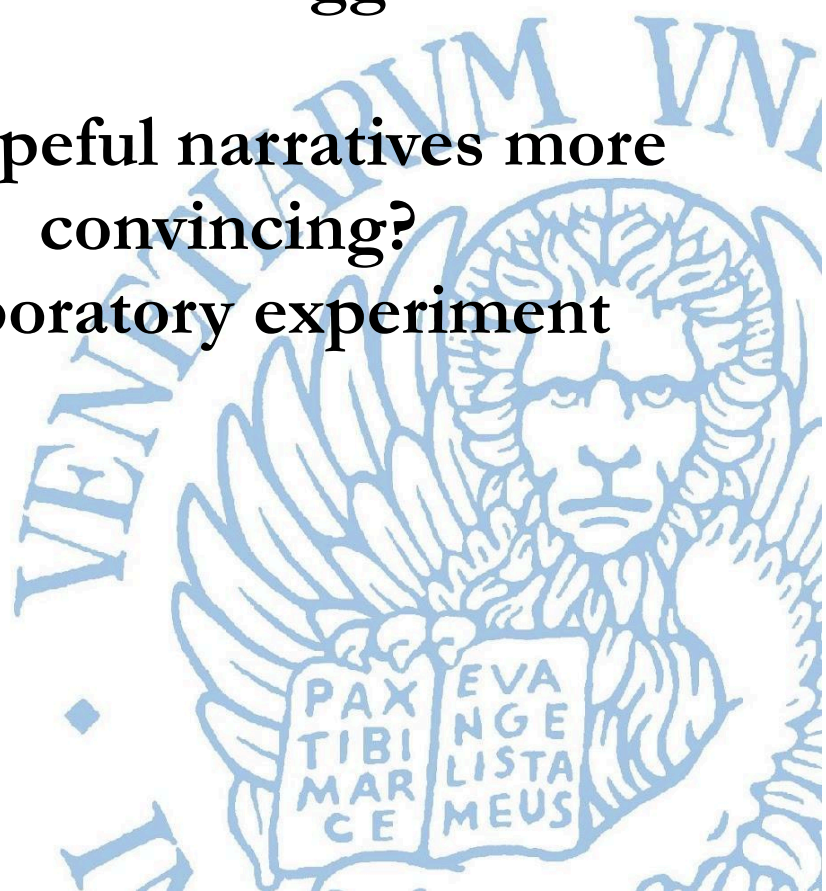
Department
of Economics

Working Paper

**Luca Corazzini
Marco Diamante
Valeria Maggian**

**Are hopeful narratives more
convincing?
A laboratory experiment**

ISSN: 1827-3580
No. 02/WP/2026





Are hopeful narratives more convincing? A laboratory experiment

Luca Corazzini

University of Milan-Bicocca

Marco Diamante

Ca' Foscari University of Venice

Valeria Maggian

Ca' Foscari University of Venice

Abstract

Assessing the causal impact of narratives on beliefs and behaviors remains an empirical challenge for social scientists, largely due to endogeneity and cultural factors. To address these limitations, we present the results of a novel, content-neutral laboratory experiment. In this experiment, participants (i) engage in a zero-sum game against a non-strategic robot, where the final outcome is determined with equal probability either by their choices or by randomness, and (ii) are exposed to either hopeful or passive narratives. These narratives differ in how ambiguous evidence is presented, suggesting whether or not participants can actively determine the final outcome of the game through their choices. Our findings reveal that, regardless of the narrative they are exposed to, participants consistently form beliefs and make choices under the illusion that they can influence the final outcomes. When provided with unambiguous evidence disproving this illusion, participants adjust their beliefs accordingly, although their choices take longer to align with these updated beliefs. Furthermore, exposure to the passive narrative reduces the inconsistency between beliefs and choices when participants mistakenly believe their choices determine the final outcome. Finally, presenting unambiguous evidence that contradicts the narrative's content increases the proportion of random and unpredictable choices.

Keywords

Narratives, polarization, illusion of control, lab experiment

JEL Codes

C91, C70, D91

Address for correspondence:

Valeria Maggian

Department of Economics

Ca' Foscari University of Venice

Cannaregio 873, Fondamenta S.Giobbe

30121 Venezia - Italy

e-mail: valeria.maggian@unive.it

This Working Paper is published under the auspices of the Department of Economics of the Ca' Foscari University of Venice. Opinions expressed herein are those of the authors and not those of the Department. The Working Paper series is designed to divulge preliminary or incomplete work, circulated to favour discussion and comments. Citation of this paper should consider its provisional character.

Are hopeful narratives more convincing?

A laboratory experiment

LUCA CORAZZINI*, MARCO DIAMANTE†, VALERIA MAGGIAN‡

Abstract

Assessing the causal impact of narratives on beliefs and behaviors remains an empirical challenge for social scientists, largely due to endogeneity and cultural factors. To address these limitations, we present the results of a novel, content-neutral laboratory experiment. In this experiment, participants (i) engage in a zero-sum game against a non-strategic robot, where the final outcome is determined with equal probability either by their choices or by randomness, and (ii) are exposed to either hopeful or passive narratives. These narratives differ in how ambiguous evidence is presented, suggesting whether or not participants can actively determine the final outcome of the game through their choices. Our findings reveal that, regardless of the narrative they are exposed to, participants consistently form beliefs and make choices under the illusion that they can influence the final outcomes. When provided with unambiguous evidence disproving this illusion, participants adjust their beliefs accordingly, although their choices take longer to align with these updated beliefs. Furthermore, exposure to the passive narrative reduces the inconsistency between beliefs and choices when participants mistakenly believe their choices determine the final outcome. Finally, presenting unambiguous evidence that contradicts the narrative's content increases the proportion of random and unpredictable choices.

Keywords: Narratives, polarization, illusion of control, lab experiment.

JEL Classification Codes: C91, C70, D91.

*Dept. Economics, Management, and Statistics; University of Milan-Bicocca (IT); . E-mail: luca.corazzini@unimib.it

†Dept. Economics; Ca' Foscari University of Venice (IT). E-mail: marco.diamante@unive.it

‡Corresponding author. Dept. Economics; Ca' Foscari University of Venice (IT). E-mail: valeria.maggian@unive.it

“Confirmations should count only if they are the result of risky predictions; that is to say, if, unenlightened by the theory in question, we should have expected an event which was incompatible with the theory, an event which would have refuted the theory” (Popper, 2014, page 36).

1 Introduction

In everyday life, people tend to interpret events, sequences of (random) facts, and vast amounts of information by constructing seemingly coherent explanations. These conjectures often lead to specific conclusions about causes or occurrences, even when they lack rigorous comparison to facts or when they are merely one among many possible explanations for a phenomenon. Without this critical evaluation, such constructs cannot be deemed scientific or even credible. Karl R. Popper, one of the 20th century’s most influential philosophers of science, argues in *“Conjectures and Refutations”* that falsifiability is the cornerstone of scientific theories. In critiquing David Hume’s approach to induction, Popper contends that scientific progress relies on prior conjectures, which are creations of the human mind¹.

Nonetheless, individuals often resort to constructing stories, better known as narratives, to explain, synthesize, and store new and unknown events in a compact, tidy, and seemingly rational form². Narratives do not necessarily need to be true; rather, to be effective, they must appear internally logical and reasonable, often blending facts, emotions, and extraneous details to create an impression on the human mind (Shiller, 2017). This potential lack of objectivity allows for the coexistence of different narratives about the same event (Eliaz and Spiegler, 2020), offering alternative reference points for individuals to form beliefs and make decisions (Tversky et al., 2000; Thaler, 2015; Thaler, 2016; Antinyan et al., 2024).

There is broad scholarly consensus on the power of narratives to shape how individuals interpret events and act accordingly (Morson and Schapiro, 2017; Shiller, 2017; Shiller, 2020). However, methodological challenges arise in empirically isolating their causal effects on beliefs and behaviors. Issues such as endogeneity and the difficulty of tracing the origin and evolution of narratives make it challenging to use real-world data to study these effects, particularly when individuals are exposed to multiple, competing narratives.

Understanding what makes one narrative more persuasive or successful than others remains an open question in the literature. An important feature influencing the success of a narrative is the role it assigns to individuals, whether it depicts them as active agents capable of influencing outcomes through their choices or as passive subjects whose actions are entirely irrelevant to determining results. In the face of an unknown event, narratives that promote a sense of agency — what we term *hopeful* narratives — may be particularly compelling. For example, consider superstition³: many people engage in superstitious practices rooted in irrational beliefs about supernatural influences because such practices provide a (false) sense of control (Case et al., 2004). Similarly, in politics, campaign messages often construct hopeful narratives that exaggerate the power of individual votes or policy measures to drive immediate change. Even in sports, fans develop

¹ “When Kant said, ‘Our intellect does not draw its laws from nature but imposes its laws upon nature’, he was right. But in thinking that these laws are necessarily true, or that we necessarily succeed in imposing them upon nature, he was wrong. Nature very often resists quite successfully, forcing us to discard our laws as refuted; but if we live, we may try again” (Popper, 2014).

² As Berger (1997) effectively puts it, “we seldom think about it, but we spend our lives immersed in narratives. Every day, we swim in a sea of stories and tales that we hear, read, listen to, or see (or some combination of all of these), from our earliest days to our deaths.”

³ Superstition is defined as “a false conception about causation or belief or practice” (www.merriam-webster.com. Retrieved 2023-08-21.)

rituals, believing they influence game outcomes.

This paper provides the first experimental evidence on whether and how, when exposed to two equally plausible and competing narratives, individuals are more inclined to believe in the hopeful story — that is, the one attributing to them the ability to influence outcomes actively.

In our laboratory experiment, participants play multiple rounds of a simple zero-sum game against a non-strategic computerized robot. Strategies are two-dimensional, involving combinations of symbols and colors. Participants are informed that, with equal probability, either the symbol or the color determines the winner of the game. Identifying the correct rule is essential for maximizing earnings. Before playing, participants are shown a subset of ambiguous outcomes consistent with both payoff rules (symbol-based and color-based). Depending on the treatment, the presentation order of these outcomes emphasizes either the symbol (*Symbol treatment*), the color (*Color treatment*), or neither (*Baseline treatment*). While participants can control the symbol, the color is randomly determined for both them and their opponent, making the symbol-narrative the more hopeful option. As in real-world contexts, where individuals and media selectively present facts to construct narratives, our experimental design exposes participants to alternative narratives — hypothetical causal models linking choices to outcomes. This controlled environment allows us to explore how narratives influence beliefs and decision-making.

Our findings reveal three key insights. First, irrespective of the narrative, participants tend to believe - and make choices that are coherent with these beliefs - that the controllable dimension (the symbol) governs the game, casting a bridge between this work and the *illusion of control* literature -Langer (1975)-. This illusion is mitigated by higher cognitive reflection and risk aversion, which help suppress intuitive but incorrect causal assumptions. Second, providing *unambiguous* evidence that the controllable dimension is irrelevant to determine earnings enables participants to update their beliefs correctly, though their choices take longer to adjust. Third, “hearing the other side of the story,” that is exposing subjects to the passive narrative in the *Color treatment*, reduces the inconsistency between stated beliefs and actual choices. Finally, in both treatments, providing unambiguous evidence that contradicts the narrative’s content leads to confusion, increasing the proportion of random and unpredictable choices.

The remainder of the paper is structured as follows: Section 2 reviews the relevant literature. Section 3 describes the experimental design and procedures. Section 4 outlines the preregistered hypotheses. Results are presented and discussed in Section 5, while Section 6 concludes.

2 Literature review

There has been increasing interest in economics regarding how narratives form and influence beliefs and behaviors (Shiller, 2017, Shiller, 2020; Morson and Schapiro, 2017). In essence, narratives are simple explanations of events that agents rely on to form an enduring understanding of reality (Shiller, 2017) and make sensible choices.

Narratives carry important elements that make them pervasive in our social life. First, “*they animate the sense-making process*” (Brooks, 1992) and provide models of understanding that agents use to organize, explain, justify, predict, and influence the course of human experience (Bruner, 1991).

Second, narratives do not necessarily need to reflect the truth, nor is it required that agents’ responses to them be fully rational. Rather, what truly matters for a narrative to be persuasive is its internal logic and

the “*perceived grain of truth*” (Bénabou et al., 2018), which results from *mixtures of fact, emotion, human interest, and other extraneous details that form an impression on the human mind*” (Shiller, 2017). In Juille and Jullien (2016), narratives are defined as stories that are used to produce consistency among the variety of choices in the different games that economic agents play simultaneously. Narrativity is described as a form of consistency, representing an alternative to rationality in economic theory.

Third, narratives enhance mental representations of reality (frames) and provide effective reference points with which we constantly engage when forming opinions and making decisions (Tversky et al., 2000; Thaler, 2015, Thaler, 2016). In this regard, Shiller (2017) emphasizes how the concept of narrative is closely linked to the *representativeness heuristic* (Kahneman and Tversky, 1973), serving as tools to form expectations about potential outcomes of a phenomenon.

Our work is inspired by Eliaz and Spiegler (2020)’s approach to narratives, which views them as tools to simplify the understanding of complex issues, proposing a story about *what causes what and what should be done*. Agents are assumed not to choose beliefs, but causal models that describe and organize the reality they observe. Hence, beliefs form by *fitting a model to historical data*, so that incorrect beliefs are merely the result of mis-specified models. In other words, narratives can be false, thus leading to distorted beliefs about the mapping from actions to consequences. On this note, several pieces of literature explain how false beliefs struggle to be eradicated, even when contradicted by evidence, leaving enduring “hard-to-die” encrustations (Esponda et al., 2024). Anderson et al. (1980) discusses something similar to narratives when referring to *causal accounts* that, by providing agents with efficient ways to organize and understand the social world, may become independent of data, contributing to the persistence of initial (incorrect) impressions.

Another possible explanation for why individuals interpret ambiguous signals based on prior beliefs is related to limited memory. Fryer et al. (2019) proposes a model in which subjects store ambiguous (hence complex and heavy) observations as simple (lighter) interpretations that are no longer ambiguous. The interpretation chosen depends on some form of prior knowledge about the phenomenon under study, which may lead them to polarize their behavior.

Previous experimental studies have primarily modeled narratives with social or moral attributes. For example, Hillenbrand and Verrina (2022) empirically tested the predictions of the model presented by Bénabou et al. (2018), which examines the production and dissemination of narratives that justify actions based on moral considerations. Their findings demonstrate that positive narratives promoting prosocial behavior increase giving in dictator games, whereas negative narratives supporting self-interested actions have varying effects depending on individuals’ social types, with selfish individuals showing an increase in their giving. Our experimental design, in contrast, is purposely aseptic. While it can easily be modified to include real-world conjectures, our goal is to provide a tool for investigating both beliefs and behavioral responses to exposure to conflicting but equally plausible narratives about reality, ruling out confounds due to subjects’ personal experiences and convictions.

The closest studies to ours are Charles and Kendall (2025) and Kendall and Oprea (2024). In the former, the authors conducted a series of controlled experiments to study how causal narratives affect subjects’ beliefs and actions, finding evidence supporting the predictions of Eliaz and Spiegler (2020). Additionally, they found that people naturally create causal narratives and have strong preferences for sharing them with others. The effects of narratives are only mitigated by factual statistical information, but they do not disappear. In the latter study, subjects were asked to form a mental model of a data-generating process, with the aim of understanding what makes some models harder to infer than others.

Compared to these works, our experimental design is intended to investigate competing narratives, with a focus on their effects on beliefs and behaviors when both suggest a potentially true causal relationship. This complements the study by [Charles and Kendall \(2025\)](#), where rationality and coherence with the narrative are always incompatible.

Our paper also contributes to the investigation of populism and demagoguery, which are currently central to policy debates. [Eliaz and Spiegler \(2020\)](#) theoretically hypothesizes that when faced with two narratives, one offering an easy fix (hence *hopeful*) and the other a rational alternative, the former guarantees popularity by conveying (false) hope for easily solving problems. Our lab-based data allows for more convincing inferences about the determinants and consequences of the success of (misleading) narratives. This, in turn, can lead to more credible policy recommendations. Indeed, demagoguery can be defined as flattering propaganda that appeals to the economic and social aspirations of the masses, relying on the illusion of easy control over political and economic phenomena.

3 Experimental Design

3.1 The basic setting

The basic setting consists of a sequential zero-sum game in which the subject acts as the second-mover and responds to a choice randomly selected by a non-strategic robot. The outcome of the game depends on the comparison between two balls: one for the robot and one for the subject. Each ball is identified by a combination of a color - either yellow, blue, or red - and a symbol - either @, % or # - resulting in nine possible combinations. While both dimensions serve to identify a ball, only the symbol can be actively chosen by the subject, making the subject's strategy space one-dimensional.

The game proceeds sequentially in three steps. First, the robot randomly selects the symbol of its ball with equal probability. After observing the robot's selection, the subject chooses the symbol of her ball. Once the subject confirms her symbol choice, the computer randomly and independently assigns a color to each of the two competing balls.

The outcome of the comparison between the two balls is determined by one of two possible payoff rules: the *Symbol Rule*, in which the winner is determined by a hierarchy of the three symbols, and the *Color Rule*, in which the winner is determined by a hierarchy of the three colors. The subject does not know which of the two payoff rules is actually used to determine the outcome of the game. The only information she has is that the actual payoff rule is randomly selected by the computer at the beginning of the experiment, with equal probability.

Depending on the payoff rule, the contest can result in one of three possible outcomes: a tie, the subject wins and the robot loses, or the subject loses and the robot wins. In the case of a tie, both the subject and the robot earn nothing. In the remaining two cases, the winner's earnings are given by 14 euros minus the cost of the corresponding ball, with the amount paid directly by the loser. As depicted in [Figure 1](#), the cost of a ball is determined by summing the cost of its color and the cost of its symbol. The color costs 6 euros if the ball is yellow, 4 euros if it is blue, and 2 euros if it is red. Similarly, the symbol costs 6 euros if it is @, 4 euros if it is %, and 2 euros if it is #.

Figure 1: Winner's earnings and cost of symbols and colors.

Gain = 14 - Cost «@» - Cost «Yellow»
 $= 14 - 6 - 6 = 2$

	Cost «Yellow»= 6	Cost «Blue»= 4	Cost «Red»= 2
Cost «@» = 6	@	@	@
	Gain= 14-6-6= 2	Gain= 14-6-4= 4	Gain= 14-6-2= 6
Cost «%» = 4	%	%	%
	Gain= 14-4-6= 4	Gain= 14-4-4= 6	Gain= 14-4-2= 8
Cost «#» = 2	#	#	#
	Gain= 14-2-6= 6	Gain= 14-2-4= 8	Gain= 14-2-2= 10

It is important to note that payoffs are structured so that the optimal symbol choice depends on the subject's beliefs about which payoff rule determines the winner. Specifically, if she believes the *Symbol Rule* applies, choosing a dominant symbol maximizes earnings (left panel of Table 1)⁴. Conversely, if the *Color Rule* applies, choosing the cheapest symbol maximizes payoffs (right panel of Table 1)⁵.

Table 1: Expected payoffs of the row player when playing any symbol, given the symbol chosen by the robot, if either the *Symbol Rule* (left panel) or *Color Rule* (right panel) is the rule governing the game.

		@	%	#
@	Mean	<u>0</u>	<u>4</u>	4
	Standard Deviation	0,00	1,73	1,73
	Minimum	0	2	2
	Maximum	0	6	6
%	Mean	-4	0	<u>6</u>
	Standard Deviation	1,73	0,00	1,73
	Minimum	-6	0	4
	Maximum	-2	0	8
#	Mean	-4	-6	0
	Standard Deviation	1,73	1,73	0,00
	Minimum	-6	-8	0
	Maximum	-2	-4	0

		@	%	#
@	Mean	0,00	-0,67	-1,33
	Standard Deviation	2,45	3,32	4,24
	Minimum	-4	-6	-8
	Maximum	4	4	4
%	Mean	0,67	0,00	-0,67
	Standard Deviation	3,32	4,12	5,00
	Minimum	-4	-6	-8
	Maximum	6	6	6
	Mean	1,33	0,67	0,00
#	Standard Deviation	4,24	5,00	5,83
	Minimum	-4	-6	-8
	Maximum	8	8	8

3.2 Information, treatments, repetition and feedbacks

Before making any choice, the subject is provided with a subsample of nine possible matches between pairs of balls, highlighting the winning/losing party and the resulting payoffs. This subsample allows the subject to infer both the hierarchy of colors and the hierarchy of symbols⁶. However, it does not reveal the actual

⁴ Choosing @ against % and # or choosing % against # guarantees the highest payoff, while choosing # likely results in a loss or a null payoff in expectation.

⁵ In this case, playing # is always optimal. Intuitively, if the winner is determined by the hierarchy of colors, choosing the lowest-cost symbol ensures the highest earnings in case of a win.

⁶ @ wins against % and #, % wins against #, if the Symbol Rule applies. Yellow wins against Blue and Red, Blue wins against Red, if the Color Rule applies.

payoff rule, as the nine matches are compatible with both the *Color Rule* and the *Symbol Rule*.

The order in which the nine matches are presented in the instructions constitutes our treatment manipulation. Specifically, our experiment includes three treatments: *BSL*, *SYM*, and *COL*. In the *BSL* treatment (see Figure A1 in the experimental instructions in Appendix A), the order of the nine matches is randomized, making it impossible to infer any specific (causal) regularity regarding whether the payoff rule is based on color or symbol⁷. Conversely, in the *SYM* (*COL*) treatment, the order of the nine matches is deliberately chosen to emphasize the *Symbol Rule* (*Color Rule*) (see Figures A2 and A3 in the experimental instructions in Appendix A), although both rules remain equally likely to have been selected by the computer.

The game described above is repeated for thirty-six rounds without feedback. However, at the end of the 9th, 18th, 27th, and 36th rounds, the subject is privately shown the outcome of an additional match randomly selected from those not included in the initial subsample. Crucially, this additional outcome is perfectly informative about the actual payoff rule⁸.

The information structure of our experiment is motivated by two key considerations. First, the absence of feedback on the subject’s choices in each round prevents strategic behavior aimed at inferring the correct payoff rule⁹. Second, the additional matches shown every nine rounds, being perfectly informative about the actual rule determining individual payoffs, allow us to assess the subjects’ (Bayesian) ability to counteract potential biases.

Our experimental design captures a fundamental characteristic of narratives: the selective connection of factual evidence to enhance the salience of a particular causal explanation among multiple possibilities. Additionally, a key advantage of our experiment is that it enables us to study the effects of narratives in a controlled, simplified, and neutral environment, thereby eliminating potential cultural and social confounding factors.

3.3 Procedures

Upon arrival at the lab, subjects were randomly assigned to a computer terminal. At the beginning of the experiment, instructions were distributed and read aloud¹⁰. Before starting the first round, subjects were asked to answer a set of control questions to ensure their understanding of the instructions. Any questions from the subjects were answered privately.

Although subjects participated in thirty-six rounds, final earnings were determined based on the outcome of the game in only one round. Specifically, at the end of the experiment, the round used to determine final earnings was randomly selected with equal probability. This choice was made to eliminate any possibility of hedging between periods, ensuring that, in each individual period, subjects had the maximum possible incentive to make decisions that would maximize their earnings.

As mentioned earlier, the probability that each of the two payoff rules would be the effective one to determine earnings was set to 50%. Specifically, subjects were told that, at the beginning of the session, they would be randomly divided into two subgroups of equal size. One subgroup was assigned to the *Color Rule*, and

⁷ The order of the nine matches in the *BSL* treatment was randomized at the experiment level.

⁸ For example, if a red @ wins against a blue #, it confirms the hierarchy of symbols and contradicts the hierarchy of colors, whereas if a yellow # wins against a red %, it confirms the hierarchy of colors and contradicts the hierarchy of symbols.

⁹ For instance, the subject could select the same symbol as the robot to determine whether a tie occurs when the randomly assigned colors differ. If a tie consistently occurs, the subject would unequivocally deduce that the *Symbol Rule* is in effect.

¹⁰ The experiment was conducted in Italian. The English version of the instructions is available in Appendix A.

the other to the *Symbol Rule*. Therefore, despite the treatment to which they were assigned, for half of the subjects in each session, the dimension emphasized by the specific order in which the subsample of nine matches was presented in the instructions did not correspond to the dimension on which the payoff rule was based (excluding the *BSL* treatment, where the subsample of matches was presented in a random order).

In the 9th, 18th, 27th, and 36th rounds of the game, before showing the subjects the outcome of a further match between balls, we elicited their belief about which payoff rule they had been assigned to. To this end, we used the *Binary Lottery Procedure* (McKelvey and Page, 1990, Schlag et al., 2013, Hossain and Okui, 2013, Harrison et al., 2014), a proper incentive-compatible mechanism for belief elicitation. In each of the four rounds in which belief elicitation was implemented, subjects were presented with two boxes positioned at the same level on the screen: in one box, they indicated the probability that the actual payoff rule was the *Color Rule*, and in the other, they indicated the probability that the actual payoff rule was the *Symbol Rule*. Both probabilities were restricted to be integers between 0 and 100, and they had to sum to 100. Each subject was informed that, at the end of the experiment, one of the four belief elicitation rounds would be randomly selected with equal probability. In the selected round, the subject would participate in one of two possible lotteries: the first (second) lottery would be implemented if the actual rule was the *Color Rule* (*Symbol Rule*). The number of tickets assigned to the subject in each lottery depended on the reported probabilities in the selected round. Specifically, conditional on the actual payoff rule, the number of lottery tickets was computed using the following two quadratic formulas:

$$\text{Tickets for the Symbol Rule} = 10000 \cdot \left[1 - \left(1 - \frac{\text{points assigned to the Symbol Rule}}{100} \right)^2 \right]$$

$$\text{Tickets for the Color Rule} = 10000 \cdot \left[1 - \left(1 - \frac{\text{points assigned to the Color Rule}}{100} \right)^2 \right]$$

Thus, depending on the stated probabilities, each subject received a number of tickets between 0 and 10,000 for each lottery. For simplicity, the tickets were numbered in ascending order, starting from 0 to the total number of tickets assigned to the subject.

Given the tickets owned, the computer randomly drew one of the 10,000 tickets, and if the subject possessed the selected ticket, they received three additional euros to be added to their overall earnings.

Although the Binary Lottery Procedure is one of the best experimental tasks to elicit probabilistic beliefs about a dichotomous variable, it can be somewhat difficult for experimental subjects to understand. To minimize the risk of confusion, the experimental program allowed subjects to repeatedly simulate the number of tickets obtained based on the points assigned to either rule.

The Binary Lottery Procedure allowed us to investigate, in an incentive-compatible way, how subjects formed beliefs about the actual payoff rule depending on the treatment they were assigned to, as well as how they responded to the new information provided at the 9th, 18th, 27th, and 36th rounds. Indeed, when both rules were equally likely to be correct, as in the first 9 rounds of the experiment, assigning 50 points to each rule maximized the probability of winning the lottery (75%). In contrast, once the subject was presented with the outcome of the first randomly selected match between balls after the 9th round, assigning all points to the hierarchy actually ruling the game became the payoff-maximizing option.

Thus, we were able to distinguish between individuals who reasoned perfectly according to Bayesian principles and those biased toward the controlled dimension. Specifically, when focusing on the first 9 rounds of the game, the former’s choices should align with those of the latter, i.e., a player who believes that the Symbol Rule determines their payoff. In fact, in Table 2, we show the expected payoff of a rational player who assigns 50% probability to either rule being the one that determines their payoff when playing any symbol chosen by the robot. Optimal actions in this case are the same as those of a subject who believes the Symbol Rule determines their payoff (as shown in the left panel of Table 1), and the beliefs elicited in the 9th round of the game allow us to distinguish between rational subjects and those biased toward the controlled dimension.

Table 2: Expected payoffs of the row player when playing any symbol, given the symbol chosen by the robot, if no information is provided about the actual rule governing the game.

		@	%	#
@	Mean	0,00	1,67	1,33
	Standard Deviation	1,68	3,51	4,17
	Minimum	-4	-6	-8
	Maximum	4	6	6
%	Mean	-1,67	0,00	2,67
	Standard Deviation	3,51	2,83	4,99
	Minimum	-6	-6	-8
	Maximum	6	6	8
#	Mean	-1,33	-2,67	0,00
	Standard Deviation	4,17	4,99	4,00
	Minimum	-6	-8	-8
	Maximum	8	8	8

The experiment took place at VERALabEx, at the Ca’ Foscari University of Venice, between December 2021 and February 2022. Each session involved between 15 and 30 subjects and lasted approximately one hour and a half. In total, 244 participants were involved in the experiments, mainly undergraduate students in Economics and Management, recruited via *ORSEE* (Greiner, 2015).

The average payment was 16.80 euros (including a show-up fee of 3 euros). At the end of the experiment, subjects completed a post-experimental questionnaire aimed at measuring their risk attitudes (Dohmen et al., 2011), ambiguity aversion (Cavatorta and Schröder, 2021), and optimism (Kemper et al., 2015). The questionnaire also included the *Cognitive Reflection Test* (Frederick, 2005, Sirota and Juanchich, 2018) and socio-demographic questions (age, previous participation in economic experiments, gender). The experiment was computerized using *zTree* (Fischbacher, 2007).

4 Preregistered hypotheses

The hypotheses discussed below were preregistered on the *OSF* platform¹¹ before conducting the experiment.

We first focus on the first 9 rounds of the game, when subjects have not yet seen any further match between balls and thus have no information about which of the two possible payoff rules they were assigned to.

¹¹ [Link to the OSF platform](#). The project received the approval of the Ethics Committee of Ca’ Foscari University of Venice on September 16th, 2021.

Specifically, in the first phase of the experiment, we expect a significantly higher proportion of subjects to choose a symbol compatible with the rule governing the game being the one suggested by the ordering of the subsample of outcomes presented in the instructions, compared to its counterpart (**Hypothesis 1**).

Moreover, we expect the effect described above to be stronger when the narrative manipulation appeals to the dimension controlled by the participant (i.e., the symbol, in the *SYM* treatment) than when the narrative suggests that the participant’s action is not relevant in defining the likelihood of winning the game (in the *COL* treatment) (**Hypothesis 2**). In the former case, the narrative gives subjects “full control” over the outcomes of the game they are about to play, adding a *hopeful* element to our framework.

In the same vein, we expect a lower proportion of choices to be compatible with either narrative when individuals are exposed to the randomly ordered subsample of outcomes (i.e., in the *BSL* treatment) than when the outcomes are ordered in one of the two possible ways. In the latter case, it may be easier for subjects to process the information and identify regularities (**Hypothesis 3**)¹².

The second part of our analysis focuses on what happens after subjects observe the first further match between balls, that is, after the 9th round of the game. Specifically, we expect that participants will be less likely to rationally update their beliefs and consequent choices when they have been previously exposed to a narrative appealing to the dimension they control, than when the narrative suggests that the participant’s action is not decisive in defining the likelihood of winning the game (**Hypothesis 4**). In general, we expect to observe an increasing number of choices compatible with the actual rule governing the game as the game is repeated (**Hypothesis 5**).

Finally, individuals’ level of cognitive reflection and ambiguity aversion should play a significant role in shaping subjects’ biases and their susceptibility to narratives. We also expect optimism to be positively related to beliefs in the narrative that appeals to the dimension controlled by the participant (**Hypothesis 6**).

5 Results

5.1 Phase 1: Rounds 1 – 9

We begin by analyzing the first phase of the experiment, which includes rounds 1 to 9. During this phase, participants had not yet observed any additional matches between balls and were unaware of which of the two possible payoff rules they had been assigned to. We examine both their choices and beliefs.

Choices. Figure 2 categorizes the choices made in phase 1 across treatments. The dark grey portion represents choices aligned with the Symbol Rule, meaning the individual selected a dominant symbol based on the symbol hierarchy (@, %, #)¹³. The light grey portion indicates choices aligned with the Color Rule,

¹² In the list of preregistered hypotheses, this hypothesis is indicated as number 4, with subsequent hypotheses following accordingly. Notably, hypothesis 3 as reported in the *OSF* document is omitted in this study, as it pertains to a treatment not conducted for the purposes of this paper. Therefore, hypotheses 3, 4, 5, and 6 in this chapter correspond to hypotheses 4, 5, 6, and 7 on the *OSF* platform.

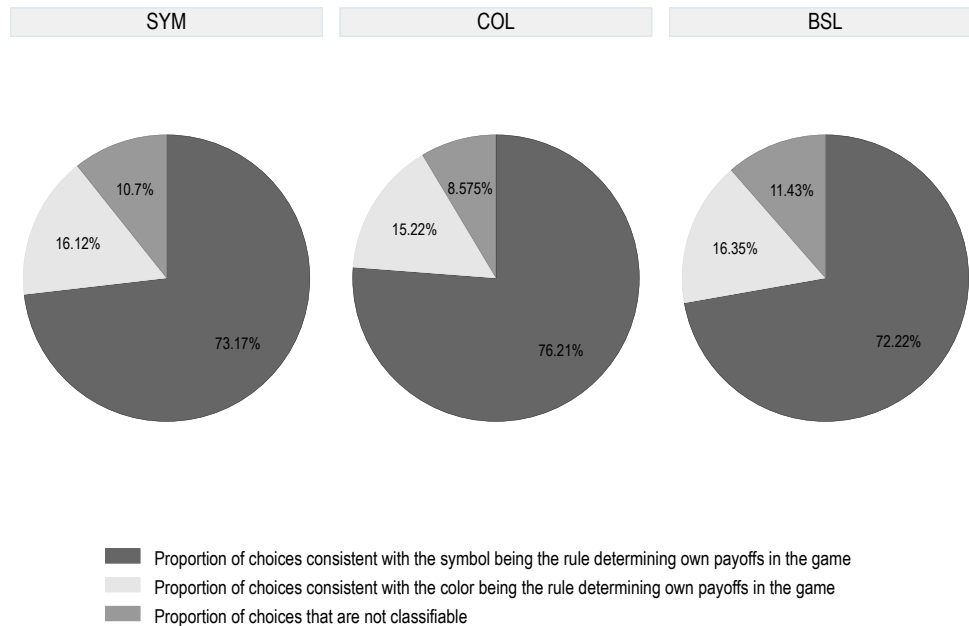
¹³ This classification implies that 4 out of 9 possible combinations are consistent with the Symbol Rule, namely @ vs @, @ vs %, and @ or % vs #. However, if participants were choosing randomly, this classification might overestimate the proportion of choices consistent with the Symbol Rule. To address this concern, in Appendix C, we adopt a more restrictive classification of choices, which yields qualitatively similar results.

in which participants selected the dominated and cheapest symbol, #. Finally, the intermediate grey portion represents choices that cannot be classified under either rationale, referred to as non-classifiable choices.

As illustrated in Figure 2, most choices align with the Symbol Rule across all treatments. A series of proportion tests provides no empirical support for Hypotheses 1 and 2, as none of the pairwise comparisons across treatments reveals significant differences in the proportion of choices consistent with either the Symbol or the Color rule¹⁴. With respect to Hypothesis 3, the proportion of non-classifiable choices is significantly higher in the *BSL* treatment than in the *COL* treatment ($p < 0.05$). This difference, however, does not emerge when comparing the *BSL* and *SYM* treatments.

Overall, this evidence suggests that most participants behaved as if the Symbol Rule—the controlled dimension—were the actual determinant of payoffs. However, it remains unclear whether choices aligned with the Symbol Rule resulted from rational behavior or from a cognitive bias toward the controlled dimension. This distinction is further explored through the analysis of beliefs elicited at the end of phase 1.

Figure 2: Proportion of choices consistent with the Symbol rule and the Color rule, by treatment - rounds 1-9.



Beliefs. Next, we examine participants’ beliefs about the payoff rule as elicited at the end of phase 1, before they observe any additional informative match between balls. This analysis allows us to assess whether choices consistent with the Symbol Rule stem from a bias toward the controlled dimension rather than from rational decision-making.

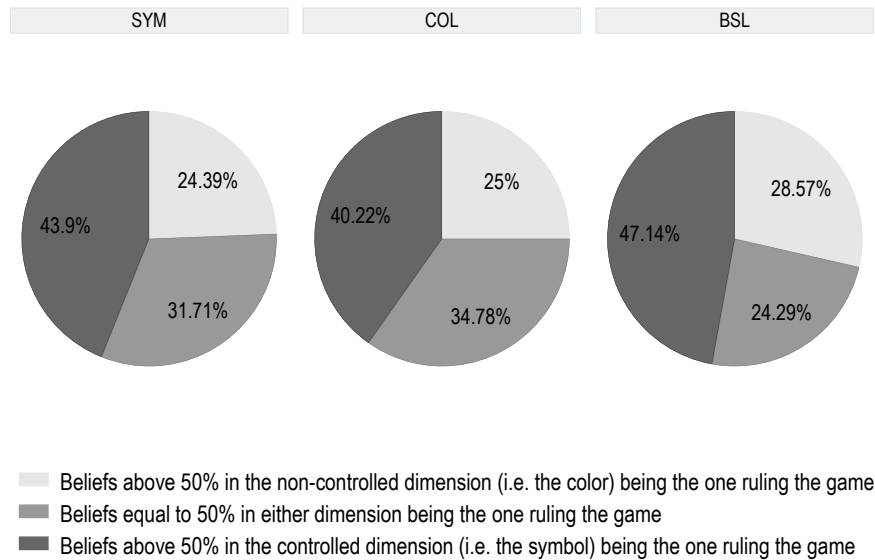
In Figure 3, the darkest portion represents individuals whose beliefs were imbalanced in favor of the Symbol Rule, the lightest portion represents individuals who believed in the Color Rule, while the intermediate grey

¹⁴ One-tailed tests. H_a : Dark slice in SYM > dark slice in COL, dark slice in SYM > dark slice in BSL, light slice in COL > light slice in SYM, light slice in COL > light slice in BSL, intermediate slice in BSL > intermediate slice in SYM, intermediate slice in BSL > intermediate slice in COL.

portion represents those who maintained a balanced (50-50) belief at the end of phase 1.

Participants were significantly more likely to favor one of the two payoff rules rather than assigning equal probability to both ($p < 0.01$, proportion test¹⁵). Among those with imbalanced beliefs, a statistically significant majority favored the Symbol Rule — the controlled dimension ($p < 0.01$ in the *SYM* treatment, $p < 0.05$ in both the *COL* treatment and the *BSL* treatment). However, there are no significant differences in the proportion of subjects with imbalanced beliefs across treatments. In other words, exposure to different narratives does not affect the tendency to form biased beliefs toward the controlled dimension¹⁶.

Figure 3: Beliefs by treatment - round 9.



We further examine whether imbalanced beliefs are associated with individual behavioral traits, as measured in the post-experimental questionnaire, including cognitive reflection, ambiguity aversion, and optimism. Specifically, in the top panel of Table 3 we first categorize participants into quartiles depending on their absolute deviation from the rational beliefs, namely assigning equal probability to both payoff rules. In the lower section of Table 3, we present the average deviation within each quartile, considering only beliefs biased toward the controlled dimension. In both cases, the first quartile consists of individuals whose deviation from the rational belief is zero, meaning they are fully rational subjects. Second, in Table 4, we estimate a set of ordered logit models for the probability that an individual belongs to a given quartile.

¹⁵ Unless otherwise stated, all tests are two-tailed.

¹⁶ Nevertheless, our narrative manipulation affects the *intensity* of the biased beliefs. Specifically, relative to the 50% benchmark, beliefs are more strongly imbalanced in the *SYM* treatment (95% confidence interval: 50.72 – 59.63) than in both the *COL* treatment (48.57 – 57.13) and the *BSL* treatment (49.43 – 60.71).

Table 3: Deviation of stated beliefs from the (50-50) beliefs benchmark

	Mean	Observations
First quartile	0	75
Second quartile	9.02	52
Third quartile	20.27	56
Fourth quartile	36.90	61

	Mean (only for values ≥ 0)	Observations
First quartile	0	75
Second quartile	8.96	28
Third quartile	20.05	36
Fourth quartile	35.12	42

Table 4: Deviations of stated beliefs from the rational (50-50) benchmark in phase 1.

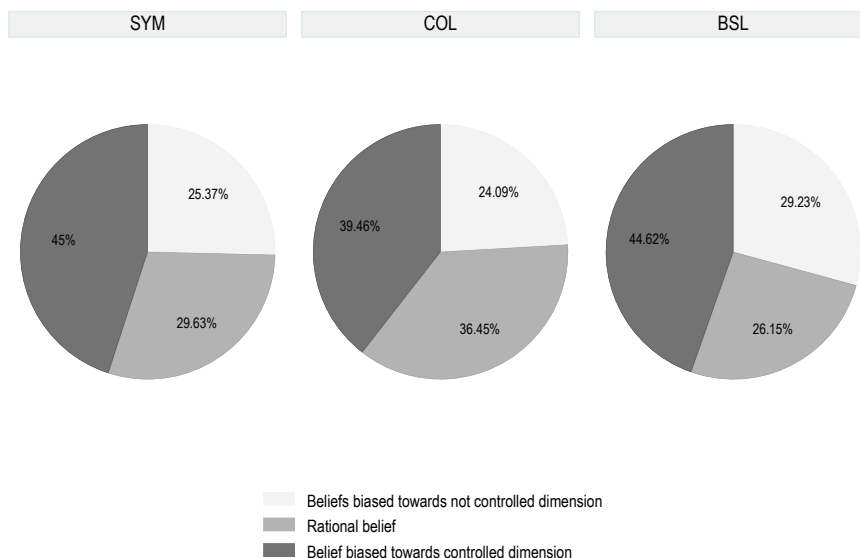
	(1)	(2)	(3)	(4)
	All	All	Symbol-biased beliefs	Symbol-biased beliefs
SYM Treatment	-0.515*	-0.419	-0.459	-0.193
	(0.285)	(0.295)	(0.346)	(0.372)
COL Treatment	-0.498*	-0.532*	-0.553	-0.481
	(0.296)	(0.321)	(0.347)	(0.382)
CRT		-0.390***		-0.440***
		(0.097)		(0.118)
Risk		0.073***		0.080***
		(0.025)		(0.029)
Ambiguity		-0.002		-0.002
		(0.004)		(0.005)
Optimism		-0.070		-0.028
		(0.077)		(0.096)
Obs.	244	244	181	181
ll	-333.9	-321.9	-236.4	-224.7
Wald $-\chi^2$	3.918	27.705	2.794	26.938
$p > \chi^2$	0.141	0.000	0.247	0.000

Note: This table reports estimates from ordered logit models (robust standard errors in parentheses). Coefficients should be interpreted in terms of direction and statistical significance rather than magnitude, since they are expressed in log-odds. The dependent variable takes an integer value from 1 to 4, corresponding to the quartiles described in Table 3 (classification reported in the top panel for columns (1) and (2), and in the bottom panel for columns (3) and (4)). *COL* and *SYM* are treatment dummies (*BSL* serves as the reference category). *CRT* measures participants' performance in the Cognitive Reflection Test (Frederick, 2005; Sirota and Juanchich, 2018) and ranges from 0 to 3, based on the number of correct answers. *Ambiguity* captures ambiguity aversion based on the Ellsberg urn questions in the post-experimental questionnaire (Cavatorta and Schröder, 2021). *Risk aversion* and *Optimism* are measured according to Dohmen et al. (2011) and Kemper et al. (2015), respectively. Significance levels are denoted as follows: * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

We observe that our narrative manipulations do not significantly affect the probability to state biased beliefs: the coefficients of *SYM* and *COL* are only marginally significant in column (1) and in columns (1) and (2), respectively. However, the coefficient of *CRT* is significant in columns (2) and (4), confirming the role of cognitive ability in reducing biases in beliefs: the higher an individual’s cognitive reflection score, the lower the likelihood of stating imbalanced beliefs, as predicted by Hypothesis 6. Similarly, risk-taking is positively correlated with the likelihood of deviating from rationality when forming beliefs.

Relationship between choices and beliefs. We now turn to the relationship between choices and beliefs, examining the extent to which the observed imbalance in beliefs is related to the bias in choices toward the Symbol Payoff Rule. Specifically, as previously noted¹⁷, in Figure 4 we restrict our analysis to choices that are consistent with the controlled dimension being the actual payoff rule, and we investigate whether these choices were made by participants with rational beliefs, by those who believe the Symbol Rule determines their payoff, or by individuals who display biased beliefs toward the non-controlled dimension.

Figure 4: Choices consistent with the Symbol payoff rule and beliefs, by treatment - rounds 1-9.



Note: The figure shows the percentage of choices consistent with the symbol rule, broken down by participants’ belief types elicited at the end of Phase 1. The light slice refers to participants with beliefs biased toward the non-controlled (i.e., the color) dimension, the intermediate slice to those with rational (50-50) beliefs, and the darkest slice to participants with imbalanced beliefs toward the controlled (i.e., the symbol) dimension.

In both the BSL and SYM treatments, choices consistent with the Symbol Rule being payoff-relevant are significantly more likely to be made by individuals who report imbalanced beliefs in favor of that rule, rather than by those with rational beliefs or beliefs toward the non-controlled dimension (Proportion test, $p < 0.01$). This suggests that the bias toward the controlled dimension observed in choices primarily reflects the behavior of participants with biased beliefs rather than those with rational beliefs. Additionally, in the

¹⁷ As shown in Table 1 and Table 2, in the first 9 rounds of the game, the optimal actions of a rational player who assigns a 50% probability to either rule being the one determining their payoff coincide with those of a subject who believes that the Symbol Rule determines their payoff.

COL treatment, choices compatible with the Symbol Rule are more likely to be made by subjects holding 50 – 50 beliefs than in either the BSL or the SYM treatments (Proportion test: $p < 0.05$ and $p < 0.01$, respectively). This suggests that “hearing the other side of the story” is beneficial: exposure to a narrative that contrasts with the controlled dimension makes observed behavior more consistent with rational beliefs.

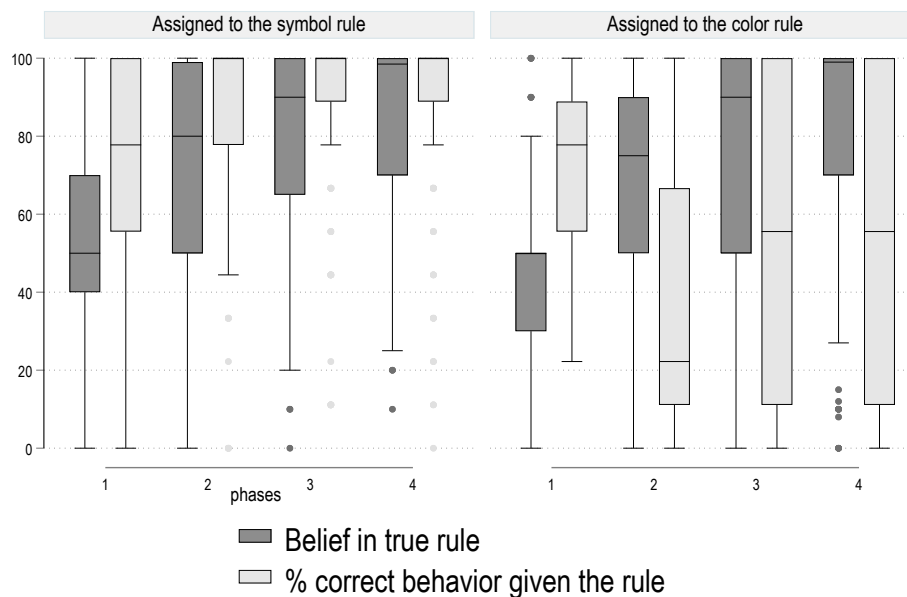
5.2 Phases 2, 3, and 4: rounds 10-36

The second part of our analysis focuses on participants’ choices and beliefs from the 10th round onward, after they have received the outcome of a further match between balls, which is perfectly informative about their actual payoff rule.

Under the hypothesis of rational Bayesian updating, after observing the outcome of the additional ball match, subjects should immediately identify the correct rule used to determine their payoffs and adjust both their beliefs and choices accordingly. This implies that subjects assigned to the Symbol Rule should always choose the dominant symbol (namely @ and %, depending on the case), while subjects assigned to the Color Rule should always choose the cheapest symbol (namely #).

Since we did not find supporting evidence for hypothesis 4, namely that the treatment manipulation affects participants’ ability to rationally update beliefs and choices (see also Appendix B), we conduct the following analysis by pooling data from all three treatments¹⁸.

Figure 5: Distributions of beliefs (dark grey boxplots) and choices (light grey boxplots) consistent with the actual rule individuals were assigned to.



Note: The left panel refers to subjects assigned to the Symbol Rule, the right panel to those assigned to the Color Rule.

Figure 5 reports the distributions of individuals’ beliefs in the rule actually determining payoffs (dark grey

¹⁸ Figure B1 in Appendix B reports the same graphs for each treatment.

boxplots) and that of the proportion of choices aligning with it (light grey boxplots). The left panel refers to subjects assigned to the Symbol Rule, while the right panel refers to those assigned to the Color Rule. Absent any bias, we should observe both the dark grey and the light grey boxplots to shift towards 100% from phase 1 to the subsequent phases of the experiment.

However, while beliefs (dark grey box plots) are updated almost immediately in the expected direction between phases 1 and 2, with no significant differences among subgroups of subjects assigned to the different rules (See the Two-sample Kolmogorov–Smirnov test for equality of distribution functions in the first column of Table 5), behavior (light grey box plots) takes more time to become consistent when the non-controlled dimension (the color) is the one determining the winner of the game, as shown by the light grey plots in the right panel of Figure 5¹⁹. In line with this, the distribution of the proportion of choices consistent with the actual rule to which the subject is assigned significantly changes ($p < 0.01$) when comparing the two possible rules phase by phase, as shown by the second column of Table 5.

Table 5: P-values of Two-sample Kolmogorov–Smirnov test for equality of distribution functions across rules.

	Beliefs	% Correct Choices
H_0 : Phase 2: assigned to Symbol rule = assigned to Color rule	0.597	0.000
H_0 : Phase 3: assigned to Symbol rule = assigned to Color rule	0.493	0.000
H_0 : Phase 4: assigned to Symbol rule = assigned to Color rule	0.704	0.000

We observe that subjects tend to converge toward rational beliefs and behavior, as predicted by Hypothesis 5. However, with respect to behavior, the rate of convergence varies depending on the rule to which they were assigned, being faster when the controlled dimension (the symbol) determines the game’s outcomes. To further investigate the role of cognitive skills in shaping (potentially biased) behavior, Figure 6 focuses only on individuals assigned to the Color Rule. Specifically, in Figure 6, we plot the distribution of each subject’s proportion of choices that are compatible, with the controlled dimension (i.e., the symbol, dark grey) or the non-controlled dimension (i.e., the color, light grey) being the rule governing the game, while intermediate grey corresponds to non-classifiable choices²⁰. We observe that the distributions of choices compatible with the Color Rule differ in phases 3 and 4, depending on the individual’s score in the Cognitive Reflection Test (Kolmogorov-Smirnov test: $p < 0.05$, see Table 6).

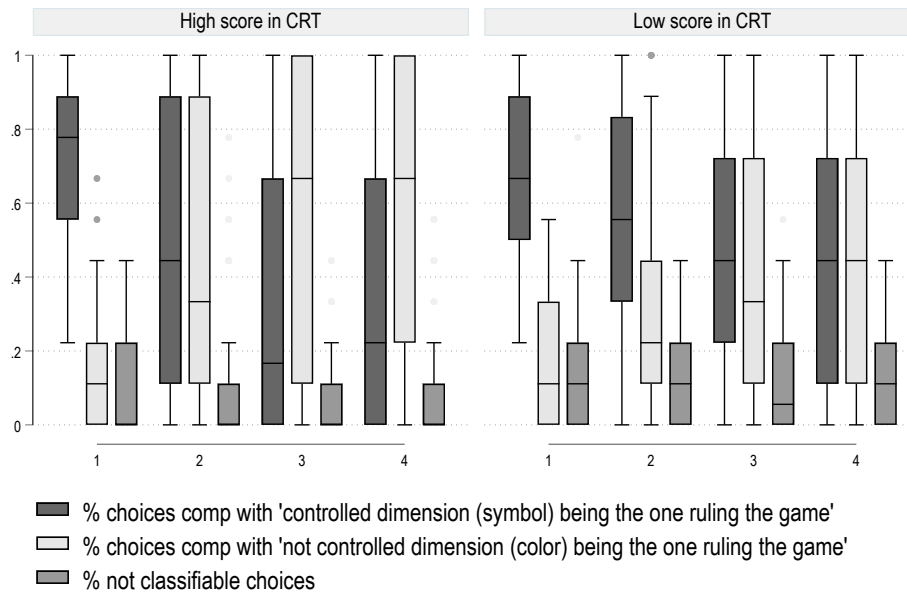
Table 6: P-values of Two-sample Kolmogorov–Smirnov test for equality of distribution functions when comparing correct choices of subjects assigned to the color rule, depending on their CRT score.

Null hypothesis	Combined K-S p-value
H_0 : Phase 2: low CRT = high CRT	0.107
H_0 : Phase 3: low CRT = high CRT	0.026
H_0 : Phase 4: low CRT = high CRT	0.017

¹⁹ The high proportion of correct choices observed in the first phase for the sub-sample of subjects randomly assigned to the Color Rule, and the subsequent decline in the second phase, should not be interpreted as misleading: as highlighted in previous sections, in the first (*blind*) phase it is optimal to behave as if the controlled dimension matters.

²⁰ Figure B2 in Appendix B reports the same graph for each treatment.

Figure 6: Individuals' behavior when the Color Rule determines outcomes, depending on their cognitive skills.



Note: The left panel refers to subjects characterized by a high score (2 or 3 correct answers) in the CRT, the right panel to those with a low score in the CRT (0 or 1 correct answers). The phases of the experiment are reported in the horizontal axis. Dark grey and light grey refers, respectively, to the proportion of choices that are compatible with the controlled dimension (i.e. symbol) and the non-controlled dimension (i.e. the color) being the one ruling the game. Intermediate grey refers to the non classifiable choices.

In general, one would expect to find correlation among the belief in a certain rule being the one determining the winner of the game and the fraction of choices consistent with it. However, we observe that, in the second phase, the Spearman's rank correlation coefficient²¹ is not significantly different from zero when considering subjects assigned to the Color Rule, with the only exception of the *COL* treatment (see the top right part of table 7), suggesting that "hearing the other side of the story", that is having been exposed to a narrative in contrast with the controlled dimension, reduces the inconsistency between beliefs and behavior when discovering that the controlled dimension is not relevant to win.

Once subjects receive additional evidence on the rule they were actually been assigned to, that is from the third phase onwards, beliefs and behavior always display a significantly different from zero positive Spearman's correlation coefficient.

²¹ The Spearman's rank correlation coefficient is computed between beliefs and choices that are consistent with actual rule determining the winner of the game.

Table 7: Spearman correlations between elicited belief and % of choices consistent with it, by phase and treatment assignment.

	Assigned to symbol rule			Assigned to color rule		
	SYM	COL	BSL	SYM	COL	BSL
Phase 2	$\rho = 0.5897$ $N = 41$ $p = 0.0000$	$\rho = 0.4165$ $N = 46$ $p = 0.0040$	$\rho = 0.3302$ $N = 35$ $p = 0.0527$	$\rho = 0.2318$ $N = 41$ $p = 0.1448$	$\rho = 0.3540$ $N = 46$ $p = 0.0158$	$\rho = 0.1732$ $N = 35$ $p = 0.3197$
Phase 3	$\rho = 0.4243$ $N = 41$ $p = 0.0057$	$\rho = 0.3047$ $N = 46$ $p = 0.0395$	$\rho = 0.5057$ $N = 35$ $p = 0.0019$	$\rho = 0.4384$ $N = 41$ $p = 0.0041$	$\rho = 0.3163$ $N = 46$ $p = 0.0323$	$\rho = 0.3576$ $N = 35$ $p = 0.0349$
Phase 4	$\rho = 0.7070$ $N = 41$ $p = 0.0000$	$\rho = 0.2833$ $N = 46$ $p = 0.0564$	$\rho = 0.4135$ $N = 35$ $p = 0.0136$	$\rho = 0.5329$ $N = 41$ $p = 0.0003$	$\rho = 0.3284$ $N = 46$ $p = 0.0258$	$\rho = 0.3917$ $N = 35$ $p = 0.0200$

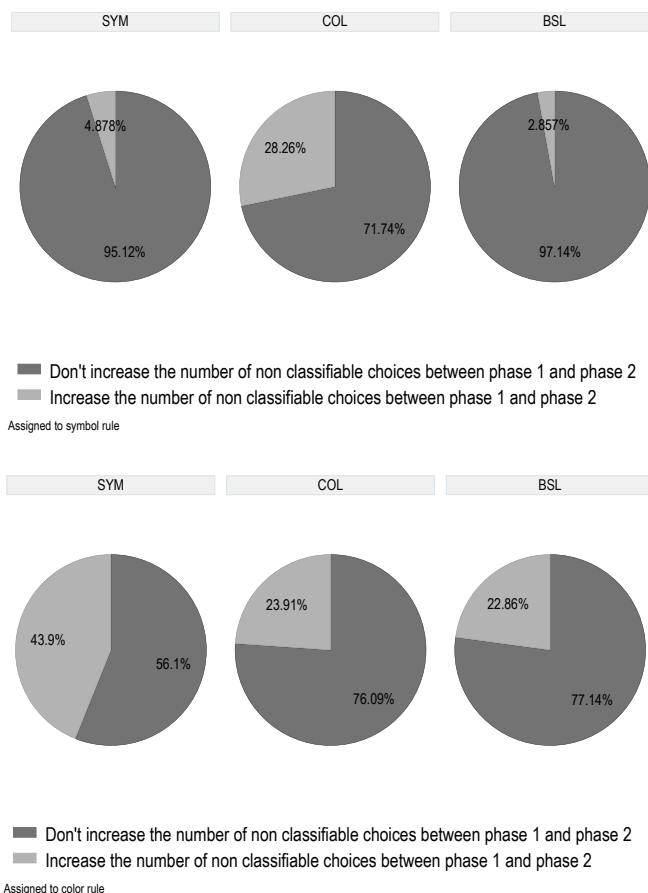
Displacement: does receiving conflicting evidence cloud one’s decision-making? To further investigate subjects’ inconsistent behavior, we analyse the proportion of non-classifiable choices. Specifically, if the hierarchy of symbols (colors) determines the winner of the game, a rational subject should play the dominant (cheapest) symbol. There are two combinations of strategies that cannot be reconciled with either hierarchy, namely playing % against @ and % against %. If such a pattern of play becomes more common after participants observe the outcome of an additional match between balls, which is perfectly informative about their actual payoff rule, we interpret this as subjects experiencing a displacement.

In the following, we investigate the role of our treatment manipulations and rule assignment in *displacing* subjects. Specifically, we consider a dummy variable whose value is equal to 1 if the proportion of each individual’s non-classifiable choices increased when moving from phase 1 to phase 2, while it is equal to 0 otherwise.

First, a two samples proportion test indicates that the probability of increasing non-classifiable choices after receiving fully informative evidence on the rule actually determining the winner of the game is significantly higher for those subjects who were randomly assigned to the Color Rule than to those who were randomly assigned to the Symbol Rule ($p < 0.01$)²². Additionally, in Figure 7, showing the fractions of *displaced* individuals for each treatment, we observe that the proportion of subjects increasing their non-classifiable choices when moving from phase 1 to phase 2 is significantly higher when subjects were exposed to either the *COL* treatment but were assigned to the Symbol Rule, or when they were exposed to the *SYM* treatment but were assigned to the Color Rule, so when instructions were suggesting a different rule than the one they were actually assigned to.

²² This is true when pooling all treatments together, but also when considering them separately, with the exception of the *COL* treatment (*SYM* treatment: $p < 0.01$, *BSL* treatment: $p < 0.05$).

Figure 7: Share of subjects (not) increasing their non-classifiable choices from phases 1 to phase 2, depending on the treatment and on the rule they were assigned to.



Note: The top panel refers to subjects assigned to the *Symbol Rule*, the bottom panel to those assigned to the *Color Rule*.

The regression analyses reported in Tables 8 and 9 consider as dependent variable the above-described dummy, taking value one if the proportion of non-classifiable choices increases when moving from phase 1 to phase 2 and taking value zero otherwise. While in Table 8 we only consider subjects assigned to the Symbol Rule, in Table 9 we only consider subjects assigned to the Color Rule.

We observe that in both Table 8 and Table 9, the coefficients of the narrative manipulations that contrast with the effective rule, as revealed by the information received in round 9, are positive and significant: the difference between the way things were “told” at the beginning and the truth revealed during the experiment, that is by the further match between balls shown to subjects, creates displacement. This, in turn, hampers participants’ ability to correctly process the information and to anticipate its implications for future behavior. Additionally, it’s noteworthy that individuals randomly assigned to the Color Rule exhibit a greater likelihood of displacement if they held a biased belief favoring the Symbol Rule at the end of the first phase compared to subjects expressing rational beliefs (see Table 9).

The coefficients of the treatment variables are no more significant if subsequent phases are considered, so that this “displacement” effect captures what happens when narratives are confuted by factual evidence for the first time (i.e. when moving from phase 1 to phase 2).

Table 8: Regression results for the probability of being displaced between phases 1 and 2, only considering subjects assigned to the Symbol Rule.

	(1)	(2)	(3)	(4)
	LPM	LPM with controls	Probit no contr	Probit with contr
SYM Treatment	0.027 (0.045)	0.029 (0.051)	0.019 (0.041)	0.018 (0.042)
COL Treatment	0.259*** (0.074)	0.260*** (0.080)	0.258*** (0.072)	0.252*** (0.073)
First belief biased twds Symbol	0.060 (0.063)	0.011 (0.081)	0.056 (0.058)	0.033 (0.071)
First belief biased twds Color	0.117 (0.083)	0.061 (0.084)	0.113 (0.072)	0.063 (0.072)
CRT		-0.035 (0.026)		-0.037 (0.024)
Risk		3.54×10^{-4} (0.007)		0.002 (0.006)
Ambiguity		0.002 (0.001)		0.001 (0.001)
Optimism		-0.024 (0.020)		-0.025 (0.015)
Constant	-0.036 (0.053)	-0.123 (0.248)		
N	122	122	122	122
ll	-31.481	-28.543	-38.689	-36.093
$Wald - \chi^2$			17.856	24.066
$p > \chi^2$	0.010	0.032	0.001	0.002

Robust standard errors in parentheses

This table analyzes the probability of increasing the proportion of non-classifiable choices between phases 1 and 2, following the informative outcome of a match between balls in round 9 shown to subjects, for those randomly assigned to the Symbol Rule. Column (1) reports the results of a linear probability model with robust standard errors, without controls. Column (2) reports results of LPM with controls, column (3) presents marginal effects from a Probit model without controls. Column (4) reports marginal effects from a Probit model with controls. The controls include the Cognitive Reflection Test (CRT) score (Frederick, 2005; Sirota and Juanchich, 2018), which ranges from 0 to 3 depending on the number of correct answers provided by each subject; Ambiguity Aversion as measured by Ellsberg urn (following Cavatorta and Schröder, 2021), risk aversion (Dohmen et al., 2011) and the degree of self declared Optimism (Kemper et al., 2015).

The statistics for the marginal effects of the probit models (N, ll, $Wald - \chi^2$, $p > \chi^2$) are derived from the respective probit models.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 9: Regression results for the probability of being displaced between phases 1 and 2, only considering subjects assigned to the Color Rule.

	(1)	(2)	(3)	(4)
	LPM	LPM with controls	Probit no contr	Probit with contr
SYM Treatment	0.223** (0.110)	0.215* (0.110)	0.212** (0.105)	0.203* (0.104)
COL Treatment	0.015 (0.098)	0.002 (0.099)	-0.005 (0.092)	-0.020 (0.091)
First belief biased twds Symbol	0.224** (0.089)	0.235** (0.092)	0.225*** (0.085)	0.237*** (0.087)
First belief biased twds Color	0.097 (0.111)	0.089 (0.113)	0.101 (0.105)	0.094 (0.104)
CRT		0.011 (0.038)		0.012 (0.037)
Risk		-0.008 (0.009)		-0.008 (0.008)
Ambiguity		-2.40×10^{-4} (0.002)		-1.06×10^{-4} (0.002)
Optimism		-0.009 (0.026)		-0.010 (0.025)
Constant	0.105 (0.101)	0.183 (0.378)		
N	122	122	122	122
ll	-72.527	-71.902	-69.262	-68.599
$Wald - \chi^2$			12.667	13.997
$p > \chi^2$	0.008	0.056	0.013	0.082

Robust standard errors in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This table analyzes the probability of increasing the proportion of non-classifiable choices between phases 1 and 2, following the informative outcome of a match between balls in round 9 shown to subjects, for those randomly assigned to the Color Rule. Column (1) reports the results of a linear probability model with robust standard errors, without controls. Column (2) reports results of LPM with controls, column (3) presents marginal effects from a Probit model without controls. Column (4) reports marginal effects from a Probit model with controls. The controls include the Cognitive Reflection Test (CRT) score (Frederick, 2005; Sirota and Juanchich, 2018), which ranges from 0 to 3 depending on the number of correct answers provided by each subject; Ambiguity Aversion as measured by Ellsberg urn (following Cavatorta and Schröder, 2021), risk aversion (Dohmen et al., 2011) and the degree of self declared Optimism (Kemper et al., 2015).

The statistics for the marginal effects of the probit models (N, ll, $Wald - \chi^2$, $p > \chi^2$) are derived from the respective probit models.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

6 Conclusions

Every day, individuals encounter controversial phenomena. Narratives provide a framework for interpreting (random) events, acting as a sense-making tool. The way ambiguous evidence is presented, by fitting facts to support a conjecture when multiple interpretations are possible, creates space for conflicting and competing narratives to emerge.

The deliberately neutral and aseptic experimental design we employ in this study, involving a zero-sum game played against a non-strategic robot with bi-dimensional strategies (combinations of symbols and colors), allows for the investigation of individuals' reactions to narratives when personal or emotional factors are absent. Specifically, before playing the zero-sum game, all subjects are presented with a subset of possible outcomes. This evidence is intentionally ambiguous, in that it is consistent with both possible payoff rules—namely, the color-based and the symbol-based rules. In the instructions, the order in which potential outcomes are presented in the subset varies depending on the treatment. This variation enables the detection of regularities, whether in terms of symbols, colors, or neither (in the baseline condition). These regularities are used to convey two distinct narratives, suggesting a causal relationship between symbols/colors and outcomes: a hopeful one, in which participants can influence the final result through their choices, and a passive one, in which outcomes are independent of participants' actions.

Our results are threefold. First, subjects exhibit biased beliefs and choices. When informed that two rules are equally likely to determine the winner of the game, they expect the rule they have control over to be the one actually governing the game. This suggests that when uncertain about the causes of an ambiguous phenomenon, people prefer to interpret reality in a way that gives them control over the situation, even if it is detrimental to their earnings.

Second, after participants receive full information about the rule that actually determines the winner of the game — that is, once any ambiguity about the reality they face is ruled out — beliefs are updated correctly, whereas individuals' biased behavior toward the controlled dimension persists.

Finally, we find that initially exposing individuals to a passive narrative—especially when they mistakenly believe their decisions ultimately determine the final outcome of the game—helps reduce the inconsistency between their beliefs and choices. However, there is also evidence that providing subjects with unambiguous information that contradicts the initial narrative's content leads to a higher proportion of random and inexplicable choices.

Our study is particularly relevant in the context of information and political consensus mechanisms: in a world characterized by uncertain cause-effect relationships, where reality is framed through narratives that tend to reinforce the salience of certain mechanisms over others, allowing individuals to believe they have control over their outcomes biases their perceptions. In this context, low cognitive ability emerges as an important determinant of this tendency.

We believe there are many promising directions for future work. Our design is purposely meant to rule out confounds due to subjects' personal experiences and convictions, but it can easily be modified to include real-world conjectures. Additionally, in our experiment, people are *forced* to learn the actual rule governing the game. It would be interesting to analyse their responses in settings where they can choose whether to receive information or where the source of such information varies, being either coming from *relevant others*, or from scientific evidence, respectively focusing on the role of social connections and trust in science in affecting what makes one narrative more persuasive or successful than others.

Acknowledgements

We acknowledge VERA - Venice centre in Economic and Risk Analytics for public policies - at the Department of Economics, Ca' Foscari University of Venice, for funding the experiment. We thank the participants at the 6th and 8th Behavioral and Experimental Economics Network (BEEN) Workshop, respectively held in Bologna and in Rome, as well as participants in the 2022 VERA Experimental and Behavioral Economics Workshop in Venice, 2022 "Bounded Rationality: Theory and Experiments" workshop at The Coller School of Management - Tel Aviv University, 2023 EEA-ESEM congress at the Barcelona School of Economics, XXXV SIEP Conference at the University of Verona and in the 2024 ASSET conference at Ca' Foscari University of Venice for useful comments.

This research project was approved by the Ethics Committee of the Ca' Foscari University of Venice on its meeting on 16 September 2021.

Data linking

The original dataset and the replication package are available at the following link: <https://data.mendeley.com/datasets/6f967y2jyk/1>

Bibliography

- Anderson, C., Lepper, M., and Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology*, 39:1037–1049.
- Antinyan, A., Bassetti, T., Corazzini, L., and Pavesi, F. (2024). Narratives and opinion polarization: a survey experiment. *Scientific Reports*, 14(19732).
- Berger, A. A. (1997). *Narratives in popular culture, media, and everyday life*. Sage.
- Brooks, P. (1992). *Reading for the plot: Design and intention in narrative*. Harvard University Press.
- Bruner, J. (1991). The narrative construction of reality. *Critical Inquiry*, 18(1):1–21.
- Bénabou, R., Falk, A., and Tirole, J. (2018). Narratives, imperatives, and moral reasoning. Working Paper 24798, National Bureau of Economic Research.
- Case, T. I., Fitness, J., Cairns, D. R., and Stevenson, R. J. (2004). Coping with uncertainty: Superstitious strategies and secondary control. *Journal of Applied Social Psychology*, 34(4):848–871.
- Cavatorta, E. and Schröder, D. (2021). Survey module to measure ambiguity preferences: Module and ambiguity preference score. *Available at SSRN 3315413*.
- Charles, C. and Kendall, C. W. (2025). Causal narratives. *SSRN Electronic Journal*. Available at SSRN: <https://ssrn.com/abstract=4669371>
Posted December 28, 2023; Last revised May 30, 2025.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.
- Eliasz, K. and Spiegel, R. (2020). A model of competing narratives. *American Economic Review*, 110(12):3786–3816.
- Esponda, I., Vespa, E., and Yuksel, S. (2024). Mental models and learning: The case of base-rate neglect. *American Economic Review*, 114(3):752–782.
- Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42.
- Fryer, Roland G. J., Harms, P., and Jackson, M. O. (2019). Updating beliefs with ambiguous evidence: Implications for polarization. *Journal of the European Economic Association*, 17:1470–1501.
- Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.
- Harrison, G. W., Martínez-Correa, J., and Swarthout, J. T. (2014). Eliciting subjective probabilities with binary lotteries. *Journal of Economic Behavior & Organization*, 101:128–140.
- Hillenbrand, A. and Verrina, E. (2022). The asymmetric effect of narratives on prosocial behavior. *Games and Economic Behavior*, 135:241–270.

- Hossain, T. and Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, 80(3):984–1001.
- Juille, T. and Jullien, D. (2016). Narrativity from the perspectives of economics and philosophy: Davis, ross, multipleselves models... and behavioral economics. GREDEG Working Papers 2016-19, Groupe de REcherche en Droit, Economie, Gestion (GREDEG CNRS), Université Côte d’Azur, France.
- Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological review*, 80(4):237.
- Kemper, C. J., Wassermann, M., Hoppe, A., Beierlein, C., and Rammstedt, B. (2015). Measuring dispositional optimism in large-scale studies. *European Journal of Psychological Assessment*.
- Kendall, C. and Oprea, R. (2024). On the complexity of forming mental models. *Quantitative Economics*, 15(1):175–211.
- Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32(2):311–328.
- McKelvey, R. D. and Page, T. (1990). Public and private information: An experimental study of information pooling. *Econometrica: Journal of the Econometric Society*, pages 1321–1339.
- Morson, G. S. and Schapiro, M. (2017). *Cents and sensibility: What economics can learn from the humanities*. Princeton University Press.
- Popper, K. (2014). *Conjectures and refutations: The growth of scientific knowledge*. routledge.
- Schlag, K. H. et al. (2013). Eliciting probabilities, means, medians, variances and covariances without assuming risk neutrality.
- Shiller, R. J. (2017). Narrative Economics. *American Economic Review*, 107(4):967–1004.
- Shiller, R. J. (2020). *Narrative economics: How stories go viral and drive major economic events*. Princeton University Press.
- Sirota, M. and Juanchich, M. (2018). Effect of response format on cognitive reflection: Validating a two- and four-option multiple choice question version of the cognitive reflection test. *Behavior research methods*, 50(6):2511–2522.
- Thaler, R. H. (2015). *Misbehaving: The making of behavioral economics*. WW Norton & Company.
- Thaler, R. H. (2016). Behavioral economics: Past, present, and future. *American economic review*, 106(7):1577–1600.
- Tversky, A., Kahneman, D., et al. (2000). *Choices, values, and frames*. Cambridge University Press New York.

Appendix A- Experimental instructions

In the following pages, we report the instructions for our three treatments (translated from Italian). The text in parentheses is specific for each treatment. We will refer to the Symbol and Color treatments respectively as SYM (also highlighted in red) and COL (also highlighted in blue), respectively.

Captions in italic are not shown to participants in the instructions.

WELCOME

Thank you for participating in this experiment.

You will receive an amount of 3€ for showing up on time. In addition, during the experiment you will be able to earn, based on your choices, an additional amount. Your final earnings will be paid to you in cash, in private, at the end of the experiment.

During the experiment you are not allowed to talk or communicate in any way with the other participants. If you have any doubts or questions during the experiment, press the “life jacket” symbol you see at the top right of the screen in front of you. An assistant will come to you to answer your questions in private.

The following rules are the same for all participants.

GENERAL RULES

The experiment consists of 36 rounds, and in each of them you will have to make a choice. Although you will participate in 36 rounds, your earnings will be determined on the basis of a single round. In particular, the round used for determining your earnings will be randomly drawn at the end of the experiment.

At the beginning of the experiment, you will be allocated an endowment of 10€. What you will obtain in the extracted round will be added or subtracted from your initial endowment of € 10.

What do you have to do? In each round of the experiment, you will challenge a robot in a simple game. The outcome of the game depends on the comparison between two balls, that of the robot and yours. The balls are identified by two dimensions: the symbol, @, % and #, and the color, Yellow, Blue and Red. The game is divided into three steps:

- Step 1. The robot will randomly select with equal probability the symbol of its ball among the three possible symbols: @, % and #.
- Step 2. After observing the robot’s ball symbol, you will be asked to choose the symbol of your ball among the three possible symbols: @, % and #.
- Step 3. Once the symbols have been determined, the computer will randomly and with equal probability select the color of each of the two balls, the robot’s ball and your own ball, among the three possible colors: Yellow, Blue and Red.










How are the gains determined in each round? In each round, your earnings depend on the outcome of the comparison between your ball and the ball of the robot. In particular, three possible outcomes can occur: (i) a tie, (ii) your ball is the winner (and that of the robot is defeated), (iii) your ball is defeated (and that of the robot wins).

In the event of a tie, both you and the robot will get a zero gain. In the other two cases, the winner will be awarded a gain of an amount that depends on the symbol and color of his/her ball. In addition, the gain

obtained by the winner in the round will be paid directly by the player who is defeated. The gain obtained by the winner in the round is given by € 14 minus the cost of his/her ball. The cost of the winning ball depends on its symbol and color, as indicated by the following table.

$$\text{Gain} = 14 - \text{Cost} \llcorner @ \gg - \text{Cost} \llcorner \text{Yellow} \gg$$

$$= 14 - 6 - 6 = 2$$

	Cost «Yellow»= 6	Cost «Blue»= 4	Cost «Red»= 2
Cost «@» = 6	 Gain= 14-6-6= 2	 Gain= 14-6-4= 4	 Gain= 14-6-2= 6
Cost «%» = 4	 Gain= 14-4-6= 4	 Gain= 14-4-4= 6	 Gain= 14-4-2= 8
Cost «#» = 2	 Gain= 14-2-6= 6	 Gain= 14-2-4= 8	 Gain= 14-2-2= 10

Any losses or gains will be subtracted from or added to the € 10 initial endowment.

What rule will be used to determine the outcome of the comparison between the two balls?

There are two possible rules that determine the result of the comparison between the two balls:

- The "Symbol Rule", according to which the winner between the two balls is determined exclusively on the basis of a specific hierarchy between the symbols @, % and #.
- The "Color Rule" according to which the winner between the two balls is determined exclusively on the basis of a specific hierarchy between the colors Yellow, Blue and Red.

At the beginning of the experiment, the computer will randomly assign half of the participants to the Symbol Rule and the remaining half of the participants to the Color Rule. The rule randomly assigned to the participant will remain the same for the whole duration of the experiment.

Which information will you have to make your choices? You will be provided with two information tools:

1. The table with a sub-sample of 9 possible comparisons between balls [SYM: sorted by symbol of the winning ball] [COL: sorted by color of the winning ball], with the corresponding results. In particular, for each comparison, the table shows you the winning ball and the corresponding earnings. We remind you that the gain associated with the winning ball will be paid by who, between you and the robot, has been defeated.

Figure A1: Order in which the sub-sample of 9 possible comparisons between balls is shown in the BSL treatment

[BSL]



















	wins against		How much?	14-4-6=4
	wins against		How much?	14-6-4=4
	wins against		How much?	14-6-6=2
	wins against		How much?	14-6-6=2
	wins against		How much?	14-6-6=2
	wins against		How much?	14-4-4=6
	wins against		How much?	14-6-4=4
	wins against		How much?	14-6-6=2
	wins against		How much?	14-4-6=4

Figure A2: Order in which the sub-sample of 9 possible comparisons between balls is shown in the SYM treatment

[SYM]





































	wins against		How much?	14-6-4=4
	wins against		How much?	14-6-6=2
	wins against		How much?	14-6-6=2
	wins against		How much?	14-6-6=2
	wins against		How much?	14-6-6=2
	wins against		How much?	14-6-4=4
	wins against		How much?	14-4-6=4
	wins against		How much?	14-4-4=6
	wins against		How much?	14-4-6=4

Figure A3: *Order in which the sub-sample of 9 possible comparisons between balls is shown in the COL treatment*

[COL]

	wins against		How much?	14-4-6=4
	wins against		How much?	14-6-6=2
	wins against		How much?	14-6-6=2
	wins against		How much?	14-6-6=2
	wins against		How much?	14-4-6=4
	wins against		How much?	14-6-6=2
	wins against		How much?	14-6-4=4
	wins against		How much?	14-6-4=4
	wins against		How much?	14-4-4=6

2. At the end of rounds 9, 18 and 27, the computer will show you the outcome of a further comparison between balls, compatible with the rule to which you have been assigned.

All the comparisons included in the two information tools described above are true and correct. This means that, if the comparison between balls in one of the rounds of the experiment were included among those indicated in the above sub-sample or in one of the further comparisons shown in rounds 9, 18 and 27, the outcome and earnings of the winning ball would be exactly as those shown.

Example 1

The robot randomly selects "#" as the symbol of its ball. The participant chooses "%" as the symbol of his/her ball. The colors randomly extracted by the computer for the two balls are respectively "Blue" for that of the robot and "Yellow" for that of the participant. The rule of comparison determines that, between the two balls, the % ball played by the participant. The gain associated with the winning ball is $14 - 4$ (cost "%") $- 6$ (cost "Yellow") = 4€. The participant wins 4€, the robot loses 4€.

Example 2

The robot randomly selects "@" as the symbol of its ball. The participant chooses "%" as the symbol of his/her ball. The colors randomly extracted by the computer for the two balls are respectively "Blue" for that of the robot and "Red" for that of the participant. The rule of comparison determines that, between the two balls, the @ ball played by the robot is the winner. The gain associated with the winning ball is $14 - 6$ (cost "@") $- 4$ (cost "Blue") = 4€. The participant loses 4€, the robot wins 4€.

Example 3

The robot randomly selects "#" as the symbol of its ball. The participant chooses "#" as the symbol of his/her ball. The colors extracted by the computer for the two balls are respectively "Red" for the robot and "Red" for the participant. The rule of comparison determines a draw between the two balls. Earnings are null for both players.

Guesswork about the rule you've been assigned to

In rounds 9, 18, 27 and 36, as soon as you have confirmed your choices, you will be asked to express two conjectures: how much you believe to have been assigned to the "Symbol Rule" and how much, instead, you believe you have been assigned to the "Color Rule".

Despite you will have to express your conjectures 4 times in the experiment, only those elicited in one of these rounds will be used to determine your final earnings. In particular, at the end of the experiment, we will randomly extract one out of four tickets numbered 9, 18, 27, 36, and the conjectures expressed in the selected round will be used to determine your gains.

How do you express your guesses in a round? Each of the two conjectures must be between 0 and 100, where 0 indicates that you do not consider at all likely that you have been assigned to that rule and 100 indicates that you are sure that you have been assigned to that rule. The sum of the two conjectures must be equal to 100.

How do you determine earnings based on your guesswork? Given the conjectures you expressed in the extracted round, you will participate in a lottery that, in the event of a favorable outcome, will give you an additional gain of € 3. The procedure used to determine the outcome of the lottery is such that you should always formulate accurate conjectures: the more likely you think that a specific rule is the one to which you have actually been assigned, the higher your conjecture about it must be. Specifically, the computer will assign a set of tickets (between 0 and 10 000, numbered progressively) to each of the two conjectures on the basis of the following formulas:

$$\begin{aligned} \text{Tickets for "conjecture prob. Symbol Rule"} &= 10000 * \left[1 - \left(1 - \frac{\text{conjecture prob. Symbol Rule}}{100} \right)^2 \right] \\ \text{Tickets for "conjecture prob. Color Rule"} &= 10000 * \left[1 - \left(1 - \frac{\text{conjecture prob. Color Rule}}{100} \right)^2 \right] \end{aligned}$$

Note that the higher the conjecture you express in correspondence with a certain rule, the higher the number of tickets included in the endowment assigned to it.

Between the two ticket endowments, you will participate in the lottery with the endowment associated with the rule of comparison between balls to which you were actually assigned by the computer at the beginning of the experiment.

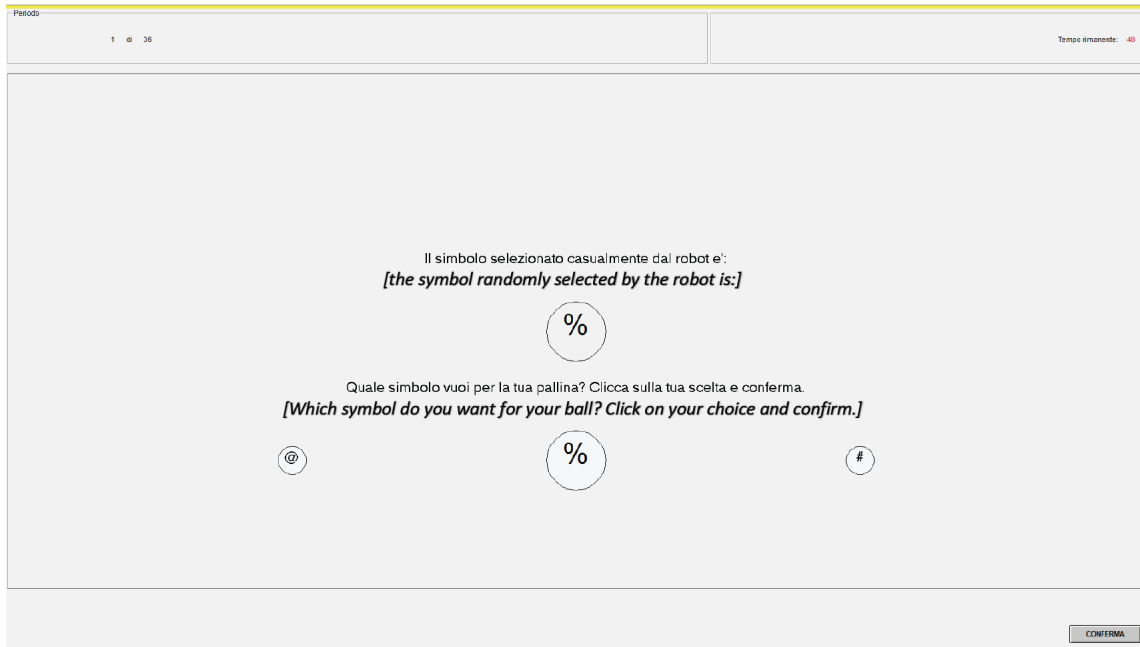
To determine the outcome of the lottery, the computer will randomly extract a ticket between 1 and 10 000 with equal probability. If the extracted ticket is included in your endowment, you will receive an additional gain of 3€. Otherwise, you will not receive any additional earnings. Note that you should always express a greater conjecture for the rule you think is most likely because, if your conjecture turns out to be correct, you will have a better chance of winning the lottery.

Example Suppose that a participant believes with probability 70 out of 100 that the rule of comparison between the balls to which it has been assigned is the "Symbol Rule" and with probability 30 out of 100 that it is the "Color Rule". In this case, based on the above formulas, the computer assigns an endowment of 9100 tickets, numbered from 1 to 9100, to the conjecture about the "Symbol Rule" and 5100 tickets, numbered from 1 to 5100, to the conjecture about the "Color Rule". Suppose that the rule to which the participant was actually assigned is the "Symbol Rule". In this case, the subject will participate in the lottery with the endowment of 9100 tickets that have been assigned to him/her for that conjecture. Let's assume that the ticket drawn from the computer to determine the outcome of the lottery is the number 7812. Since the subject owns the winning ticket, he/she gets the 3€.

Summarizing...

What do you have to do?










(i) In each round of the experiment the robot will randomly and with equal probability select the symbol of its ball among the three possible symbols: @, % and #. After observing the robot's symbol, you will be asked to choose the symbol of your ball among the three possible symbols: @, % and #.



(ii) The computer will randomly and with equal probability select the color of each ball among the three possible colors: Yellow, Blue and Red.

How are the earnings of the round determined?

(iii) Your gains in each round depend on the outcome of the comparison between your ball and the ball of the robot. In the event of a tie, both you and the robot will get a zero gain. In cases where one of the two balls is the winner, the gain obtained by the winner in the round will be paid directly by the one who, between you and the robot, was defeated. The gain obtained by the winner is given by € 14 minus the cost of his/her ball determined on the basis of the following table.

	Cost «Yellow»= 6	Cost «Blue»= 4	Cost «Red»= 2
Cost «@» = 6	 Gain= 14-6-6= 2	 Gain= 14-6-4= 4	 Gain= 14-6-2= 6
Cost «%» = 4	 Gain= 14-4-6= 4	 Gain= 14-4-4= 6	 Gain= 14-4-2= 8
Cost «#» = 2	 Gain= 14-2-6= 6	 Gain= 14-2-4= 8	 Gain= 14-2-2= 10

Gain = 14 - Cost «@» - Cost «Yellow»
 = 14 - 6 - 6 = **2**

Any losses or gains will be subtracted from or added to the € 10 initial endowment.



















What rule will be used to determine the outcome of the comparison between balls?

(iv) There are two possible rules that determine the result of the comparison between the two balls: (i) the "Symbol Rule", according to which the winner between the two balls is determined exclusively on the basis of a specific hierarchy between the symbols @, % and #; (ii) the "Color Rule", according to which the winner between the two balls is determined exclusively on the basis of a specific hierarchy between the colors Yellow, Blue and Red. Throughout the experiment, based on a random computer extraction made at the beginning, half of the participants will be assigned to the "Symbol Rule" and the remaining half to the "Color Rule".

(v) In order to support you in your choices, you will be provided with two information tools:

1. The table with a sub-sample of 9 possible comparisons between balls [SYM: sorted by symbol of the winning ball] [COL: sorted by color of the winning ball], with the corresponding results. For each comparison, the table shows you the winning ball and the corresponding earnings.

[BSL]

	wins against		How much?	14-4-6=4
	wins against		How much?	14-6-4=4
	wins against		How much?	14-6-6=2
	wins against		How much?	14-6-6=2
	wins against		How much?	14-6-6=2
	wins against		How much?	14-4-4=6
	wins against		How much?	14-6-4=4
	wins against		How much?	14-6-6=2
	wins against		How much?	14-4-6=4

[SYM]

@	wins against	%	How much?	14-6-4=4
@	wins against	%	How much?	14-6-6=2
@	wins against	%	How much?	14-6-6=2
@	wins against	#	How much?	14-6-6=2
@	wins against	#	How much?	14-6-6=2
@	wins against	#	How much?	14-6-4=4
%	wins against	#	How much?	14-4-6=4
%	wins against	#	How much?	14-4-4=6
%	wins against	#	How much?	14-4-6=4

[COL]

%	wins against	#	How much?	14-4-6=4
@	wins against	%	How much?	14-6-6=2
@	wins against	#	How much?	14-6-6=2
@	wins against	#	How much?	14-6-6=2
%	wins against	#	How much?	14-4-6=4
@	wins against	%	How much?	14-6-6=2
@	wins against	%	How much?	14-6-4=4
@	wins against	#	How much?	14-6-4=4
%	wins against	#	How much?	14-4-4=6

2. At the end of rounds 9, 18 and 27, the computer will show you the outcome of a further comparison between balls, compatible with the rule to which you have been assigned.

All the comparisons included in the two information tools described above are true and correct. This means that, if the comparison between balls in one of the rounds of the experiment were included among those indicated in the above sub-sample or in one of the further comparisons shown in rounds 9, 18 and 27, the outcome and earnings of the winning ball would be exactly as shown.

Guesswork about the rule you've been assigned to

(vi) In rounds 9, 18, 27 and 36, as soon as you have confirmed your choices, you will be asked to express

your conjectures, from 0 to 100, about the probability of having been assigned to the "Symbol Rule" and the "Color Rule".

Periodo 9 di 36 Tempo rimanente: 41

Congiecture del periodo 9
[Conjectures of period 9]

CONGIETTURE SULLA REGOLA CHE DETERMINA L'ESITO DEL CONFRONTO TRA PALLINE
[CONJECTURES ON THE RULE THAT DETERMINES THE OUTCOME OF THE COMPARISON BETWEEN BALLS]

Indica quanto ritieni probabile che la regola che determina l'esito del confronto tra palline a cui sei stato assegnato sia la "regola simbolo" e quanto ritieni probabile che sia la "regola colore".
[State how much you think it's likely that the rule determining the outcome of the comparison between balls to which you have been assigned is the "symbol rule" and how much you think it's likely it is the "color rule"]

"REGOLA SIMBOLO" [SYMBOL RULE]

La regola a cui sei stato assegnato e' quella che si basa esclusivamente sulla gerarchia di simboli
[The rule to which you have been assigned is the one basing exclusively on the hierarchy of symbols]

70

BIGLIETTI [TICKETS]

9100

SIMULA BIGLIETTI

[SIMULATE TICKETS]

"REGOLA COLORE" [COLOR RULE]

La regola a cui sei stato assegnato e' quella che si basa esclusivamente sulla gerarchia di colori
[The rule to which you have been assigned is the one basing exclusively on the hierarchy of colors]

30

BIGLIETTI [TICKETS]

5100

OK

Your guesswork will result into two possible ticket endowments for a lottery. The higher the conjecture you express in correspondence with a certain rule, the higher the number of tickets included in the endowment assigned to it. Between the two ticket endowments, the one with which you will participate in the lottery is the endowment associated with the rule of comparison between balls to which you were actually assigned by the computer at the beginning of the experiment.

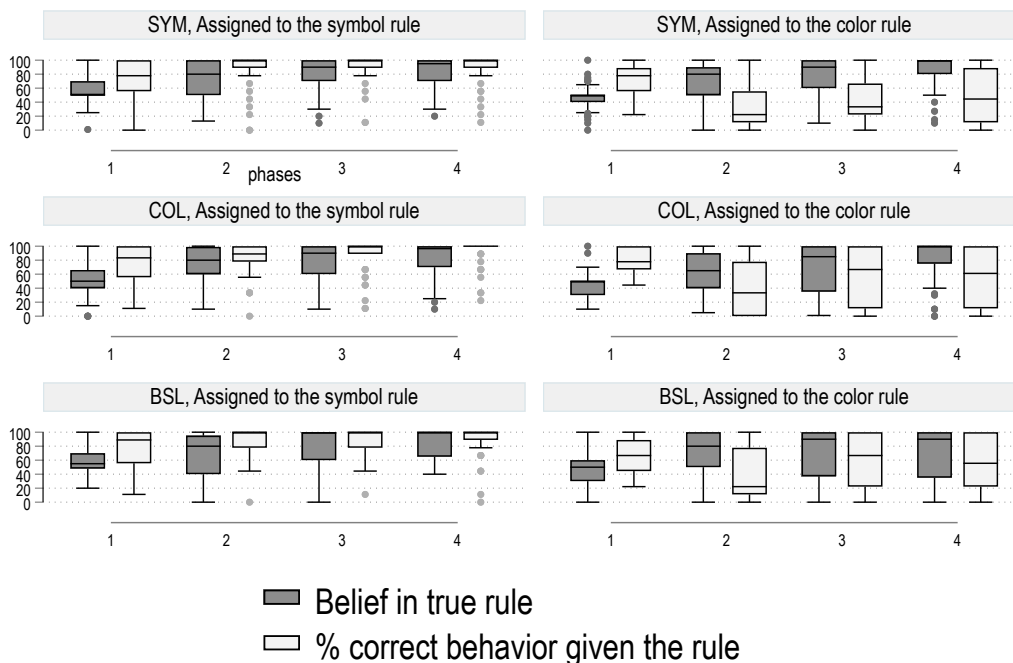
To determine the outcome of the lottery, the computer will randomly extract a ticket between 1 and 10 000 with equal probability. If the extracted ticket is included in your endowment, you will receive an additional gain of 3€, otherwise, you will not receive any additional earnings. This procedure is such that it is always advisable to formulate accurate conjectures: the more likely you think that a specific rule is the one to which you have actually been assigned, the higher your conjecture about it will have to be.

(vii) To determine your final gain, at the end of the experiment only one round will be randomly drawn among the 36 in which you had to make a choice. Similarly, a single round between 9, 18, 27 and 36 will be randomly drawn for the payment of your guesswork.

Appendix B - Additional analyses

Figure B1 replicates Figure 5 in sub-section 5.2: it illustrates the distributions of correct beliefs, depicted as dark grey boxplots, and the proportion of choices aligning with the assigned rule, depicted as light grey boxplots, in each treatment. In each row, the left (right) panel refers subjects assigned to the Symbol (Color) Rule.

Figure B1: Distributions of correct beliefs (dark grey boxplots) and proportion of choices consistent with the rule subjects are assigned to (light grey boxplots), by treatment.



Note: For each row, the left panel refers to subjects assigned to the Symbol Rule, the right panel to those assigned to the Color Rule.

We conduct a series of two-sample Kolmogorov–Smirnov tests to compare the empirical distributions of participants’ stated beliefs within each experimental phase, treatment by treatment, conditional on the rule to which they were assigned. Specifically, we test whether, for subjects in the SYM treatment in Phase 2, the distribution of beliefs differs depending on whether they learned that the payoff-relevant rule was the Symbol rule or the Color rule. Equivalently, we assess whether the two dark-gray boxplots for Phase 2 in the top panels of the figure—each corresponding to a different rule—represent statistically distinct distributions or overlapping ones. We repeat this procedure for all phases and treatments. We then perform the same analysis for the distributions represented by the light-gray boxplots, which capture the proportion of correct choices relative to the rule that actually determined payoffs.

Table B1: Two-sample Kolmogorov–Smirnov tests: p-values for comparisons of the distributions of beliefs across rules, by treatment and phase.

	Phase 2	Phase 3	Phase 4
SYM	0.990	0.990	1.000
COL	0.087	0.490	0.829
BSL	0.976	0.683	0.115

Overall, the Kolmogorov–Smirnov tests indicate that distributions of beliefs do not differ significantly across treatments in any phase, with only a marginal exception in the second phase of the COL treatment (see Table B1).

Table B2: Two-sample Kolmogorov–Smirnov tests: Combined K-S p-values for comparisons of the distributions of % of correct choices across rules, by treatment and phase.

	Phase 2	Phase 3	Phase 4
SYM	0.000	0.000	0.000
COL	0.000	0.002	0.001
BSL	0.000	0.007	0.003

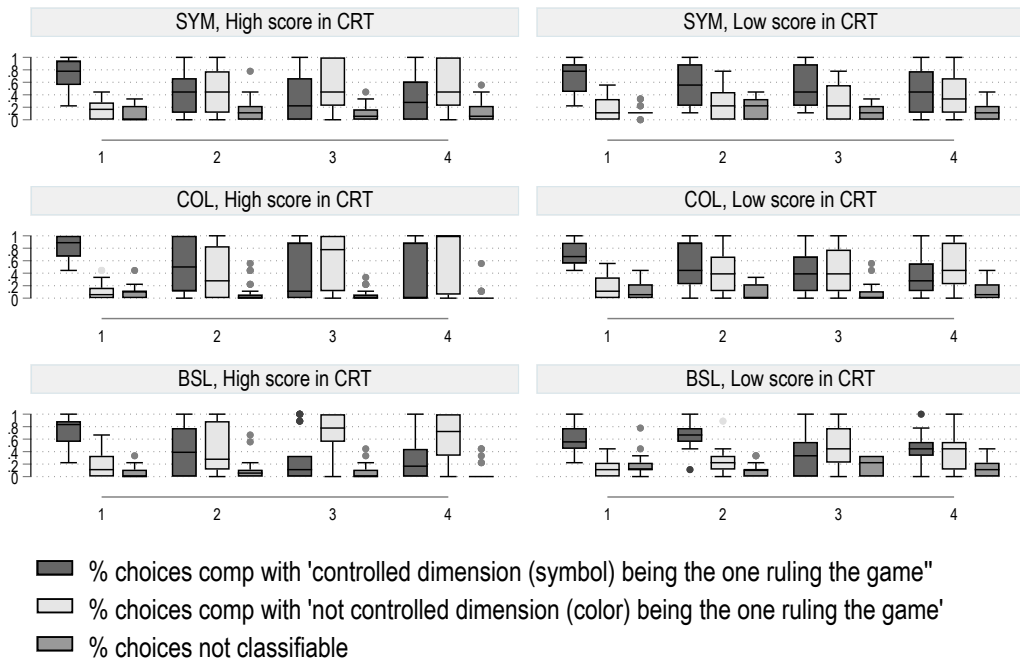
In contrast, Table B2 shows that, when comparing the distributions of the proportion of correct choices across the payoff-relevant rules—phase by phase and treatment by treatment—we consistently observe statistically significant differences ($p < 0.01$) between subjects assigned to the Symbol Rule and those assigned to the Color Rule.

Our additional analysis confirms our main results: in each treatment, individuals are shown to be able to update their beliefs in the expected direction between phases, with no significant differences depending on the rule they were assigned to. Behavior, however, takes more time, and individuals are less likely to adjust their choices when they have been informed that the non-controlled dimension, rather than the controlled one, actually determines payoffs.

Figure B2 replicates Figure 6 in section 5.2, by treatment. Specifically, as in the main text, Figure B2 only includes those individuals who were assigned to the Color Rule, as this is the case in which participants take longer to adjust their behavior correctly after receiving the additional outcome of a match between balls, perfectly revealing the actual payoff rule. Individuals are categorized according to their score on the CRT task. Dark gray indicates the proportion of choices compatible with the controlled dimension (the symbol), while light gray indicates choices compatible with the uncontrolled dimension (the color). Choices that cannot be classified are shown in intermediate gray.

Similarly to what observed when considering all treatments together, individuals with low CRT scores are slower in making correct choices than those with high CRT scores, even if this difference fails to reach statistical significance due to the restricted number of observations.

Figure B2: Individuals' behavior when the Color Rule determines outcomes, depending on their cognitive skills, by treatment.



Appendix C - Alternative categorization of choices

The categorization of combinations of subjects’ and robots’ choices used in the main text implies that a relative majority of these combinations (4 out of 9) is compatible with the Symbol Rule (i.e., the controlled dimension matters), whereas only 3 out of 9 are compatible with the Color Rule (i.e., the non-controlled dimension matters), and 2 out of 9 are compatible with neither rule. This is illustrated in Table C1. Given this asymmetry, one might argue that if subjects were choosing completely at random, the higher frequency of Symbol Rule-compatible combinations would mechanically generate a relative majority of choices attributed to the Symbol Rule.

In order to check whether our results might be driven by this concern, we replicate the main figures and statistical tests that could be affected by it using an alternative categorization of choices that is equally logically defensible, but in which the frequencies of choices attributable to the Symbol Rule, the Color Rule, and neither rule are each exactly one third (see table C2).

Table C1 clarifies the categorization used in the main text, with participants represented as row players and the robot’s choices shown in columns. Under this classification, any dominant symbol (in the case of ’@,’ the symbol capable of tying, i.e., ’@’ itself) is considered a choice compatible with the Symbol Rule, relative to the symbol played by the robot.

Table C1: Categorization of row player choices.

	@	%	#
@	Compatible with Controlled dimension matters	Compatible with Controlled dimension matters	Compatible with Controlled dimension matters
%	Not classifiable	Not Classifiable	Compatible with controlled dimension matters
#	Compatible with Non-controlled dimension matters	Compatible with Non-controlled dimension matters	Compatible with Non-controlled dimension matters

In Table C2, we present the alternative classification of choices used in this appendix. Specifically, here we attribute a choice to the Symbol Rule only if it is the optimal choice given that the Symbol rule is in effect. In fact, only one cell changes (the top-right cell of the table). Based on Table 1, it is clear that the best response to the robot’s ”#”, if one believes that the symbol rule is actually ruling the game, is to play ”%”. Playing ”@” is dominant but not optimal. This adjustment ensures that all three scenarios are characterized by the same probability (1 out of 3). Specifically, if a bias in choices in favor of the Symbol Rule still emerges under this classification, it can no longer be attributed—even in part—to the fact that a greater number of choice combinations are compatible with that specific rule. In this way, we have eliminated a potential confounding factor.

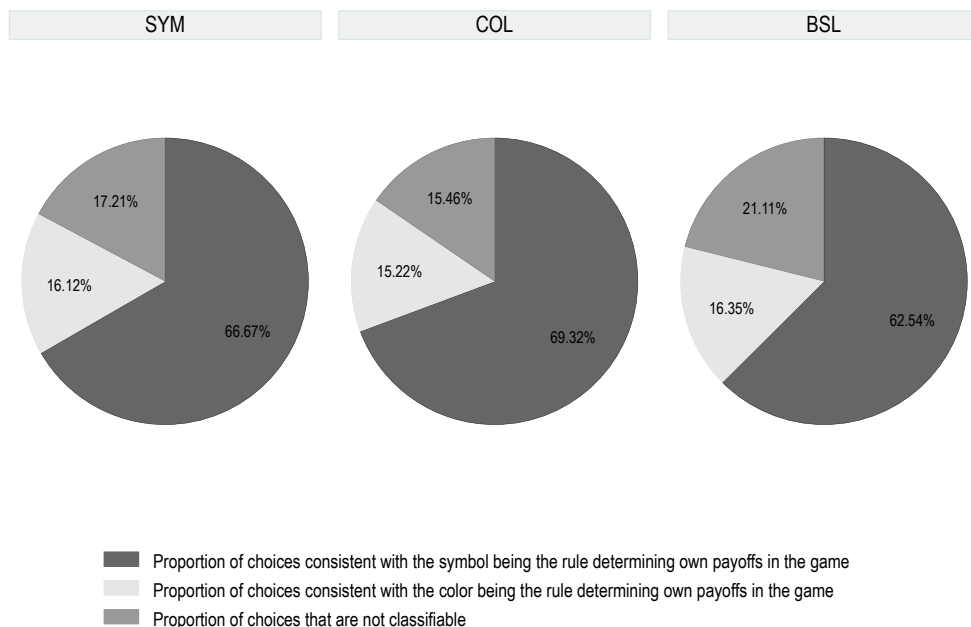
Table C2: Alternative categorization of row player choices.

	@	%	#
@	Compatible with Controlled dimension matters	Compatible with Controlled dimension matters	Non-classifiable
%	Not classifiable	Non-Classifiable	Compatible with controlled dimension matters
#	Compatible with Non-controlled dimension matters	Compatible with Non-controlled dimension matters	Compatible with Non-controlled dimension matters

In the following, we replicate Figures 2, 4, 5 and 6 in the main text using the alternative classification of choices provided in Table C2 as a robustness check. Specifically, our goal is to ensure that our results are not driven by the categorization in Table C1, in which a majority of choices are defined as compatible with the Symbol Rule simply because, relatively speaking, more combinations of actions can be attributed

to it if subjects played randomly. However, we do not replicate the analysis of individuals' displacement using the alternative categorization, as its outcome would be misleading. Specifically, under this alternative classification, individuals could be labeled as “displaced” due to an increase in choices of the type '@' vs '#', which can still be rationalized within the framework of the Symbol Rule. When investigating displacement, our goal is indeed to rely solely on choices that are truly irreconcilable with either of the two rules.

Figure C1: Consistency of choices with the Symbol and the Color Rules in the three treatments - rounds 1-9 (Replication of Figure 2 in the main text, using an alternative categorization of choices).

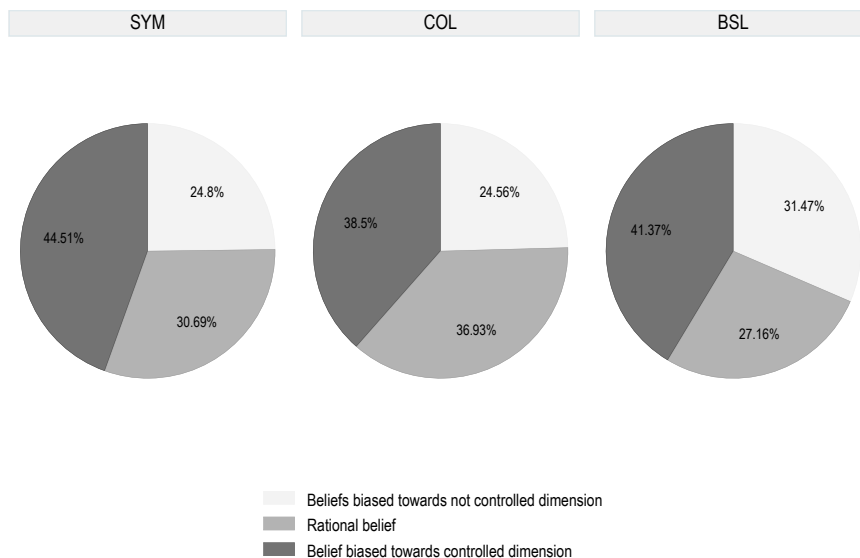


Note: Each panel refers to a treatment. The darkest slice refers to choices compatible with the *Symbol Rule*, the lightest to choices compatible with the *Color Rule*, the intermediate-grey slice of the pie refers to the fraction of non-classifiable choices.

In Figure C1 we observe that our main results are confirmed. Specifically, the proportion of choices consistent with the controlled dimension (i.e. the symbol) being the actual payoff determinant in phase 1 is not significantly different when considering pairwise comparisons of treatments²³, with just one marginally significant result when considering the *SYM* vs *BSL* ($p < 0.1$). Similarly, the proportion of choices consistent with the Color Rule being the actual payoff determinant in phase 1 is not higher in the *COL* treatment than in either the *SYM* treatment, or the *BSL* treatment. Conversely, the fraction of choices not attributable to any rule is greater in the *BSL* treatment than in both the *SYM* treatment ($p < 0.05$) and the *COL* treatment ($p < 0.01$). Finally, consistent with the findings reported in sub-section 5.1, we still observe a majority of choices compatible with the Symbol Rule in each treatment, even when using the alternative categorization of choices ($p < 0.01$).

²³ One-tailed tests. H_a : Dark slice in SYM > dark slice in COL, dark slice in SYM > dark slice in BSL, light slice in COL > light slice in SYM, light slice in COL > light slice in BSL, intermediate slice in BSL > intermediate slice in SYM, intermediate slice in BSL > intermediate slice in COL.

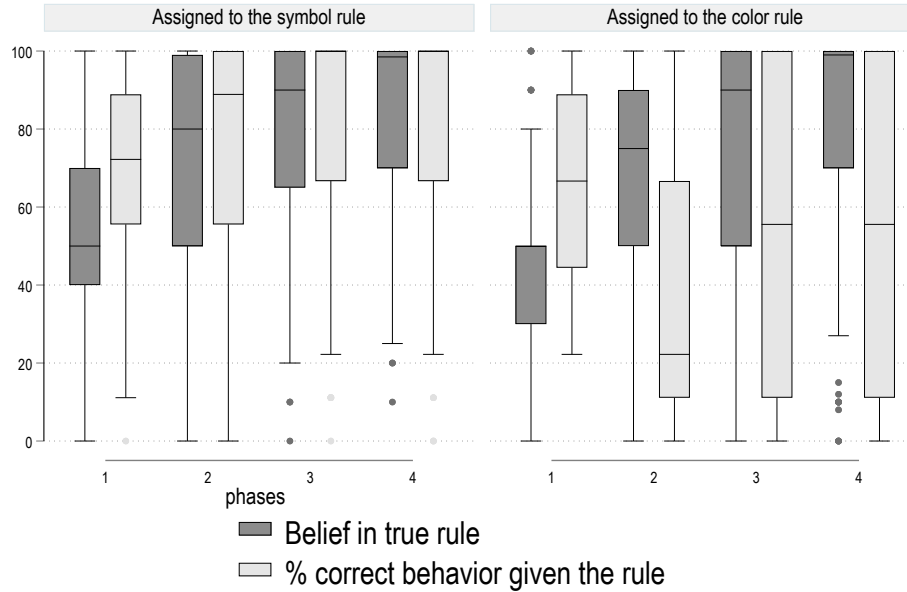
Figure C2: Choices consistent with the Symbol payoff rule and stated beliefs in the three treatments - rounds 1-9 (Replication of Figure 4 in the main text, using an alternative categorization of choices).



Note: Each panel refers to a treatment. The light slice of the pie refers to choices compatible with the *Symbol Rule* expressed by people with beliefs biased towards the non-controlled dimension, the intermediate one to those expressed by subjects with rational beliefs (50-50), the darkest one to choices compatible with the *Symbol Rule* of participants with beliefs unbalanced towards the controlled dimension.

Figure C2 focuses on the first phase of the experiment and analyzes, for each treatment, the proportion of choices consistent with the controlled dimension being the payoff rule, depending on participants' beliefs. Light gray indicates individuals whose beliefs are biased toward the uncontrolled dimension; intermediate gray corresponds to participants with rational beliefs; and dark gray to those whose beliefs are biased toward the controlled dimension. The pattern is unchanged relative to 4. In all treatments, with the exception of the *COL* one, a significantly higher proportion of choices consistent with the symbol being the payoff-relevant rule is attributed to individuals with imbalanced beliefs favoring the controlled dimension (proportion test, $p < 0.01$). Notably, in the *SYM* treatment, this share (the darkest segment in Figure C2) is larger than in the *COL* treatment (proportion test, $p < 0.05$), putting into evidence the role of our narrative manipulation in reinforcing the observed bias. Moreover, the share of choices consistent with the symbol being the payoff-relevant rule made by individuals stating rational beliefs (represented by the intermediate grey segment) is larger in the *COL* treatment than both in the *SYM* treatment and the *BSL* treatment (proportion tests: $p < 0.05$ and $p < 0.01$, respectively). Again, "hearing the other side of the story", that is having been exposed to a narrative in contrast with the controlled dimension tends to rationalize beliefs.

Figure C3: Distributions of beliefs (dark grey boxplots) and choices (light grey boxplots) consistent with the actual rule individuals were assigned to. Distributions of correct beliefs given the rule (dark grey boxplots) and proportion of choices consistent with the rule subjects are assigned to (light grey boxplots). (Replication of Figure 5 in the main text, using an alternative categorization of choices).



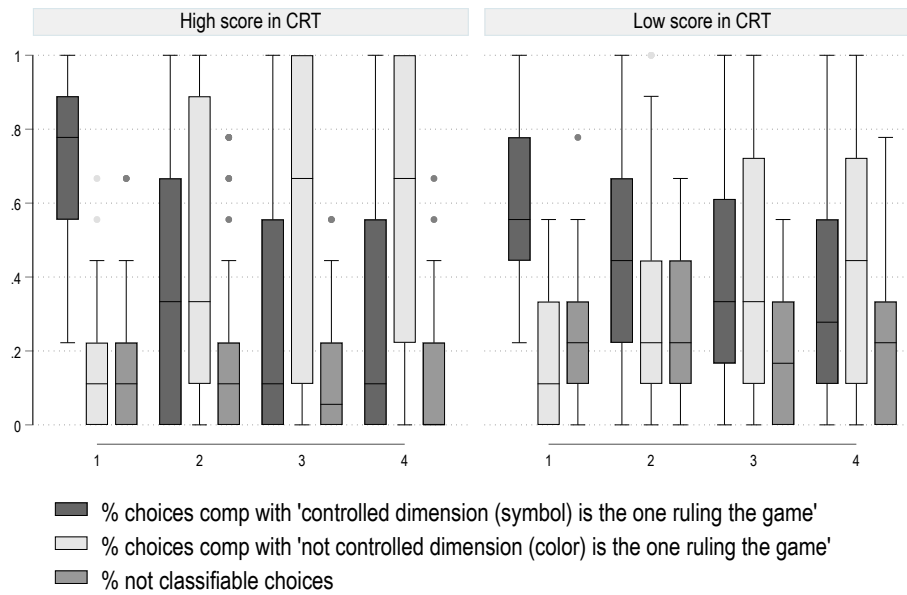
Note: The left panel refers to subjects assigned to the Symbol Rule, the right panel to those assigned to the Color Rule.

In Figure C3, nothing changes with respect to Figure 5. Specifically, the distribution of the percentage of correct choices, given the rule each subject is assigned to, changes significantly ($p < 0.01$) when comparing the two possible rules phase by phase, as shown by Table C3.

Table C3: P-values of Two-sample Kolmogorov–Smirnov test for equality of distribution functions. Replication of the second column of table 5 with alternative categorization of choices.

	% Correct Choices
H0: Phase 2: assigned to Symbol rule = assigned to Color rule	0.000
H0: Phase 3: assigned to Symbol rule = assigned to Color rule	0.000
H0: Phase 4: assigned to Symbol rule = assigned to Color rule	0.000

Figure C4: Individuals' behavior when the Color Rule determines outcomes, depending on their cognitive skills (Replication of Figure 6 in the main text, using an alternative categorization of choices).



Note: The left panel refers to subjects with a high CRT score (2 or 3 correct answers), while the right panel refers to those with a low CRT score (0 or 1 correct answers). The phases of the experiment are reported in the horizontal axis. Dark grey and light grey refers, respectively, to the proportion of choices that are compatible with the controlled dimension (i.e. symbol) and the non-controlled dimension (i.e. the color) being the one ruling the game. Intermediate grey refers to the non-classifiable choices.

The Kolmogorov–Smirnov test confirms the pattern observed in Figure 6: the distributions of the percentage of choices compatible with the Color Rule differ in phases 3 and 4 depending on the individual's CRT category ($p < 0.05$).