6-23-2020

Strategic Problems with Risky Prospects

Alessandro Sontuoso
Chapman University, sontuoso@chapman.edu

Cristina Bicchieri
University of Pennsylvania, cb36@sas.upenn.edu

Alexander Funcke
University of Pennsylvania, funcke@0z.se

Einav Hart
University of Pennsylvania, einavi@gmail.com

Strategic Problems with Risky Prospects

Comments

ESI Working Paper 20-23

Strategic Problems with Risky Prospects

Alessandro Sontuoso,[a,b] ***Cristina Bicchieri***,[b] ***Alexander Funcke***,[b] **and** ***Einav Hart***[b]

June 23, 2020

Abstract. We study "hypothetical reasoning" in games where the impact of *risky prospects* (chance moves with commonly-known conditional probabilities) is compounded by *strategic uncertainty*. We embed such games in an environment that permits us to verify if risk-taking behavior is affected by information that reduces the extent of strategic uncertainty. We then test some implications of expected utility theory, while making minimal assumptions about individuals' (risk or ambiguity) attitudes. Results indicate an effect of the information on behavior: this effect is triggered in some cases by a belief-revision about others' actions, and in other cases by a reversal in risk preferences.

KEYWORDS: Hypothetical reasoning; Strategic uncertainty; Belief revision; Risk; Complexity.

JEL Classification Numbers: C72, C92, D81, D83.

Acknowledgments: We thank Jayson Dorsett for managing the experimental lab and for programming the code used in the experiment.

Contact: [a] *Smith Institute for Political Economy and Philosophy*, Chapman University, 303 W. Palm Ave., Orange, CA, 92866. [b] *Center for Social Norms and Behavioral Dynamics*, University of Pennsylvania, 249 S. 36th St., Philadelphia, PA, 19104. E-mail: sontuoso@sas.upenn.edu; cb36@sas.upenn.edu; funcke@0z.se; einavi@gmail.com.

I. Introduction

The assumption that an agent is a (subjective) expected utility maximizer underlies most economic models. In combination with expected utility theory, a standard assumption is that an agent processes new information by revising her beliefs according to some rule (e.g., Bayesian or otherwise). We refer to a theory that incorporates these two broad assumptions as an *updated* expected utility (henceforth UEU) framework. Such a model is the most common paradigm for examining choice behavior in both non-strategic and strategic (i.e., interactive) domains. Common violations of UEU are typically attributed to two broad types of causes: agents fail to consistently update beliefs,[1] or exhibit non-standard (risk or ambiguity) attitudes.[2]

Here we investigate an additional source of deviations from UEU in strategic problems, namely, individuals' difficulty with "hypothetical reasoning" (i.e., the act of considering alternative strategically-relevant contingencies). We note that hypothetical reasoning may be especially challenging when contingencies depend on multiple unknowns: in particular, when outcomes depend both on *risky prospects* (i.e., chance moves with commonly-known conditional probabilities) and on *strategic uncertainty* (i.e., ambiguity arising from uncertainty about other players' actions and beliefs). Still, in everyday life we often need to engage in complex hypothetical reasoning: examples range from competing agents' investment decisions in the face of random economic shocks, to public health decisions on the verge of a possible epidemic.

In this paper, we study multi-player games where the impact of risky prospects is compounded by the presence of strategic uncertainty. As shall be discussed, we embed such games in an experimental environment that allows for the revision of *beliefs about other players*: this has the goal of varying (the extent of) strategic uncertainty, while leaving risky prospects

[1] Agents may exhibit unresponsiveness to information "inconsistent" with their priors, either because they ignore a signal due to inattention, or because they are characterized by some sticky or otherwise biased belief-updating process. For example, an agent may assess the probability of events by "availability" (i.e., the ease with which instances of events come to mind; Tversky and Kahneman, 1973). When processing signals drawn from a sample, one may misinterpret information as if it confirmed previous hypotheses (Rabin and Schrag, 1999), or misunderstand the relationship between sample proportions and the population from which they are drawn (Benjamin, Rabin, and Raymond, 2016). Another common heuristic involves some form of reinforcement, where one is more likely to choose acts associated with good outcomes one experienced in the past (Erev and Roth, 1998). Other belief-updating processes allowing for limited rationality are surveyed by Rabin (2013).

[2] For an instance of non-standard *risk* attitudes, see Kahneman and Tversky (1979) and Kőszegi and Rabin (2006), among others; for a survey, see O'Donoghue and Somerville (2018). For a decision-theoretic account of non-neutral *ambiguity* attitudes, see Gilboa and Schmeidler (1989); recent models are reviewed by Machina and Siniscalchi (2013). For evidence of ambiguity aversion in lotteries, see Baillon, Huang, Selim, and Wakker (2018). For an early game-theoretic account of equilibrium play in games with ambiguity, see Eichberger and Kelsey (2000); for experimental evidence, see Eichberger, Kelsey, and Schipper (2008) and Heinemann, Nagel, and Ockenfels (2009).

unchanged. We then investigate the relationship between (updated) beliefs and best-responses, and test some aggregate-level implications of expected utility theory. Notably, our analysis identifies a violation of UEU – in the form of a reversal in risk preferences – which is attributable *neither* to failures in belief updating *nor* to non-standard (risk or ambiguity) attitudes. We impute this violation to individuals' difficulty with hypothetical reasoning.

We stress that several types of interactions simultaneously present risky *and* uncertain prospects.[3] For example, think of an individual's choice of whether or not to get a vaccine: by vaccinating oneself one contributes at a cost to oneself to a particular public good, namely, herd immunity. Typically, in this type of problems, one's payoff from "free riding" on others' efforts depends not only on the population-level behavior (like in any threshold public-goods game), but also on a move by nature. Generalizing, here a simple model could posit that *if* too many people free ride – and so herd immunity is not reached – then there may or may not be a random change in the ecology (such as a random shock with a commonly-known conditional probability). In other words, *conditional on* the population's behavior, here a chance move ultimately determines if there is an epidemic outbreak or if instead the pathogen dies out.

In order to study individuals' difficulties with hypothetical reasoning, we propose an experimental game that embodies some of the key elements exhibited by the vaccination example (a "threshold game with risky prospects"). Specifically, we consider a choice problem in which outcomes depend on the population's behavior and on a move by nature, as follows. *If* the population-level frequency of a socially-undesirable action exceeds a certain threshold (e.g., if too many individuals in the population free ride), then a random shock will occur with a commonly-known conditional probability (e.g., an epidemic may occur). *If* instead the threshold is not met (e.g., enough individuals vaccinate), then outcomes will solely depend on players' choices and a random shock will not occur. As we shall discuss, the game presents players with two strategic options (vaccinate or free ride), in addition to an exit option with a sure payoff that is independent of others or nature (e.g., quarantine, the action any ambiguity- or risk-averse individual should strictly prefer). In short, best-responses depend on individuals' beliefs about the population's behavior, as well as on their risk preferences.

[3] In the case of risky prospects, the objective probability of random events is available; whereas in the case of uncertain prospects, it is not. For example – in the domain of strategic interaction – "strategic uncertainty" may occur in games with complete information and *multiple equilibria*: think of any coordination games, threshold public-goods games, etc. (Brandenburger, 1996).

Our "main treatment" embeds the above game in a network structure whereby each participant can observe the choices made by a randomly-generated *sample of fellow participants*: this has the purpose of somewhat lessening the extent of strategic uncertainty. More precisely, each participant is provided with summary information ("feedback") about some actions taken by participants that have been randomly selected to be her "neighbors" in a network.[4] (In this regard, we stress that the network structure of our design is merely a device for generating noisy information; i.e., our experiment is not designed to analyze learning in relation to structural properties of a network, as subjects had no knowledge of their position in the network or of the network structure itself.) That said, recalling that a random shock depends on the (entire) population's behavior, a subject may use the information about her neighbors' actions so as to revise the probability that the threshold has been passed at the population level.[5] We note that the goal of such a manipulation is to check if, under minimal assumptions, we can rule out failures in belief updating as an explanation for violations of UEU in our game.

Our between-subjects design also includes a "control treatment" which is the same as our main treatment, except that subjects receive no feedback. (In addition to the choice task, both the main and control treatments directly elicit subjects' beliefs about the population's behavior.) Given this, our experimental design permits us to test whether an effect of the feedback on behavior is entirely due to a *belief revision* (about the others' actions), or may be attributed to a *reversal in risk preferences*.

Our data reveal an effect of the feedback on risk-taking behavior. Furthermore, our econometric analysis determines exactly when such an effect is explained by one mechanism rather than the other (i.e., a belief revision or a reversal in risk preferences). We begin by commenting on the first explanation. In brief, the data show that the feedback does trigger a *belief revision* (which is compatible with Bayesian reasoning, on average). To that end, our analysis isolates cases in which different feedback implies different posterior beliefs, and hence

[4] Subjects were informed that each participant in the room was connected to some others at random, such that everyone was either *directly* or *indirectly* connected to everyone else. Participants directly connected to one another are referred to as "neighbors". Note that subjects had no knowledge of the specific network structure. Also note that the network structure of our design only determines what feedback is passed on to subjects, and *not* their incentives (i.e., payoffs depend on the population's behavior, not just on the neighbors' behavior).

[5] The generic nature of the feedback implies that it may well be interpreted differently by different subjects: this means that the experimenter cannot verify if subjects perform (Bayesian) belief updating at the individual level. However, as we will show, our design can verify if subjects' behavior is compatible with Bayesian updating at the aggregate level.

different best-responses. In those cases, we find evidence of behavioral changes that are attributable to a rational belief revision.

Notably, our design also allows us to identify cases in which changes in behavior may not be rationalized as the effect of a belief revision on a subject's expected utility.[6] We attribute such changes to a *reversal in risk preferences*, triggered by individuals' difficulties with hypothetical reasoning. To identify such a reversal, we isolate observations for which a rational belief update (resulting from the feedback) should *not* imply a best-response other than the action one would normally choose in the absence of feedback. Comparing choice distributions across treatments – with and without feedback – our analysis reveals that subjects collectively take the riskless action more frequently than subjects would otherwise do in the absence of feedback, given the same stated beliefs. This suggests that the experimental manipulation (i.e., merely providing feedback about others) causes a different appreciation of the risky prospects. We ascribe this effect to some subjects' inability to deal with hypothetical reasoning in a complex environment.

In this regard, we note that previous research has shown that the complexity of a choice problem influences risk-taking behavior, causing deviations from expected-value maximization in lottery choice tasks (e.g., Huck and Weizsacker, 1999). Some contributions have gone on to show that the complexity of a task affects belief updating in objective or ambiguous lotteries (Charness and Levin, 2005; Zizzo, Stolarz-Fantino, Wen, and Fantino, 2000). Other research in the domain of *non*-strategic problems has recently provided some evidence that suboptimal behavior may be due to difficulties with hypothetical reasoning (Charness and Levin, 2009; Martínez-Marquina, Niederle, and Vespa, 2018). Similarly, Esponda and Vespa (2014) propose a voting problem where players have to make some inferences under the hypothetical case that their vote is pivotal. In this clever experiment, subjects play against opponents simulated by computers: by doing so, the problem preserves the structure of a (strategic) game in terms of hypothetical contingencies, while removing any strategic uncertainty. Esponda and Vespa find that subjects make mistakes, mostly because of an inability to make inferences from hypothetical events. Finally, Levin, Peck, and Ivanov (2016) have recently found evidence of overbidding in auctions with incomplete information about the distribution of bidders' valuations, a fact that

[6] More precisely, we identify cases in which behavior changes may *not* be rationalized as the effect of a belief revision on an individual's expected utility, given either standard-selfish or "conformist" preferences.

they also ascribe to an inability to make inferences from hypothetical events (as well as to computational difficulties with Bayesian updating).

Our results complement previous findings by investigating a distinctly-different environment, that is, complete-information games with chance moves and multiple equilibria (specifically, threshold games with risky prospects), played against actual participants.[7] Furthermore, by eliciting beliefs we can directly investigate the relationship between preferences and (updated) beliefs: this allows us to test some aggregate-level implications of expected utility theory in terms of *risk-taking* behavior. In particular, for the first time we provide evidence as to how risk-taking behavior is affected by information that varies the extent of (yet does not remove) strategic uncertainty. Overall, our data show that the difficulties with hypothetical reasoning are aggravated by the strategic uncertainty inherent to a game with multiple equilibria. As the experimenter's feedback slightly attenuates one of the sources of complexity of the problem (i.e., strategic uncertainty), some subjects appear to focus on the other source of complexity, namely risk. Indeed, the data show that information about the others' behavior causes subjects to become more risk averse than they would be with no such information (on average), *given the same beliefs*. So, this means that the feedback effectively causes some subjects to come to a different "understanding" of the problem's risky contingencies.

In summary, our paper shows evidence of feedback-induced changes in behavior, which in some cases are consistent with a *belief revision* while in other cases are due to an apparent *preference reversal*. To the best of our knowledge, this paper is the first to systematically study threshold games with risky prospects, and isolate a reversal in risk preferences in such games. The remainder of the article is organized in this manner: section II presents the game, along with the experimental design and procedures; section III lays out theoretical predictions and experimental hypotheses; section IV discusses the experimental results, and section V concludes.

[7] In this regard, previous research has shown that individuals behave differently in a game played against actual participants, as opposed to a structurally-similar problem where co-players are replaced by computers (i.e., an effect that is attributed to particular risk and ambiguity attitudes; see Bohnet and Zeckhauser, 2004, and Ivanov, 2011). As we shall discuss, our design rules out non-standard (risk or ambiguity) attitudes as an explanation for deviations from expected utility theory in our data.

II. Game, experimental design, and procedures

1. *Threshold games with risky prospects*

Consider an n-player simultaneous-move game such that, for each player i, payoffs depend on a chance move θ, as well as on i's action s_i and her coplayers' actions s_{-i}. Formally, i's payoff is defined by $m_i(\theta, s)$, where $\theta \in \{H, T\}$ denotes a move by nature while $s = (s_i, s_{-i})$ compactly denotes an action profile (i.e., an n-tuple of actions), with $s_i \in \{A, B, C\}$ for each i. The move by nature is interpreted as the toss of a fair coin, resulting in either of two outcomes (HEADS or TAILS), each with a 50 percent probability. It is assumed that (without knowing the outcome of the coin toss) each player i simultaneously chooses an action; relatedly, note that the outcome of the coin toss is the same for all players.

In order to compactly define the outcomes of the game (after players have taken action), we now introduce two classes of action profiles, $\boldsymbol{\omega_0}$ and $\boldsymbol{\omega_1}$, based on whether a "threshold" $p\%$ has been met or not, as follows: we say that $s \in \boldsymbol{\omega_0}$ if *less than* $p\%$ *of all players have chosen* B; instead, we say that $s \in \boldsymbol{\omega_1}$ if $p\%$ *or more of all players have chosen* B. Given this, we say that a game is a "threshold game with risky prospects" whenever payoffs vary with θ, and $\boldsymbol{\omega_0}$, $\boldsymbol{\omega_1}$. Specifically, we assume that payoffs satisfy the following conditions:

$$\text{for any } s = (s_i^*, s_{-i}') \text{ s.t. } s \in \boldsymbol{\omega_0}, \quad m_i(H, s_i^*, s_{-i}') = m_i(T, s_i^*, s_{-i}'); \tag{1}$$

$$\text{for any } s = (s_i^*, s_{-i}') \text{ s.t. } s \in \boldsymbol{\omega_1}, \quad m_i(H, s_i^*, s_{-i}') \neq m_i(T, s_i^*, s_{-i}'); \tag{2}$$

$$\text{for any } s_{-i}, s_{-i}', \quad m_i(H, C, s_{-i}) = m_i(T, C, s_{-i}'). \tag{3}$$

Condition 1 says that payoffs do *not* vary with the coin toss at $\boldsymbol{\omega_0}$, whereas condition 2 says that they do vary with the coin toss at $\boldsymbol{\omega_1}$; lastly, condition 3 says that payoffs from C do not vary, regardless of the coin toss or the coplayers' actions.

In what follows we consider the same parameterization of the game we used in the experimental sessions, and provide some possible interpretations. (In this regard, note that the descriptive labels below – e.g., "contribute", "free ride", etc. – are reported merely for ease of illustration, but are *not* part of the game or the experimental instructions.) In short, assume $p\% = 0.4$. Next, in order to specify payoffs, suppose the coin has been tossed and all players have simultaneously chosen an action. Then, depending on the outcome of the coin toss θ, and on whether the action profile is described by either $\boldsymbol{\omega_0}$ or $\boldsymbol{\omega_1}$, one of the following alternative contingencies ("scenarios") occurs: in each case, payoffs to player i are defined as follows.

- **Scenario X.** If *less than 40%* of all players choose B (i.e., $s \in \boldsymbol{\omega_0}$), then regardless of the outcome of the coin toss θ: a player receives 0.5 payoff units if she chose A, 3 if she chose B, and 0.75 if she chose C, as indicated in Table 1 below.

- If *40% or more* of all players choose B (i.e., $s \in \boldsymbol{\omega_1}$), then two scenarios are possible:
  - **Scenario Y.** When the coin outcome θ is HEADS—a player's payoff is 1 if she chose A, -1.5 if she chose B, and 0.75 if she chose C.
  - **Scenario Z.** When the coin outcome θ is TAILS—a player's payoff is 0.5 if she chose A, 3 if she chose B, and 0.75 if she chose C.

In plain words, players choose among three actions, ranging from a very risky option "B" (with both the highest and lowest possible payoffs), to a mildly risky option "A", and finally to a riskless option "C". Note: for reasons that will be evident when we state the experimental predictions, we set this particular parameterization of the game in such a way that, at $\boldsymbol{\omega_1}$, the expected values of the payoffs from the three actions are the same (recall that the coin results in either of two outcomes, HEADS or TAILS, with a 50 percent probability).

As an application of such a game, for illustrative purposes we discuss the case of an individual's decision of whether to get a costly vaccine, on the verge of a possible epidemic outbreak. In that case, by choosing C (*exit*, e.g., self-isolation) one removes oneself from the set of susceptible individuals, which explains the scenario-invariant payoff associated with this option. On the other hand, by choosing A (*contribute*, e.g., vaccination) one contributes – at a cost to oneself – to the public good, namely herd immunity; in this case, one's utility depends on the others' decisions (which determine whether the threshold is reached) as well as on the outcome of a move by nature. By contrast, by choosing B (*free ride*, e.g., no vaccination) one negatively contributes to the threshold for herd immunity. So, one's payoff from choosing B depends on the population-level behavior and on a move by nature, which results in a 50 percent chance of an epidemic outbreak *if* too many people free ride. In summary, whenever less than 40% of individuals free ride (that is, when more than 60% of the population either get a vaccine or self-isolate), then the state of herd immunity occurs, i.e., "scenario X". On the other hand, when 40% or more free ride, then with a 50 percent chance a negative shock in the form of an epidemic will occur (i.e., "scenario Y"). When 40% or more free ride *and* a negative shock does not occur (i.e., "scenario Z"), individuals end up receiving the same payoffs as under herd immunity. A compact representation of an individual's payoffs is given below.

	A	B	C
	contribute *(e.g., vaccination)*	*free ride* *(e.g., no vaccination)*	*exit* *(e.g., self-isolation)*
scenario X (i.e., $\boldsymbol{\omega_0}$): *herd immunity*	0.5	3	0.75
scenario Y (i.e., $\boldsymbol{\omega_1}, HEADS$): *no herd immunity, epidemic outbreak*	1	-1.5	0.75
scenario Z (i.e., $\boldsymbol{\omega_1}, TAILS$): *no herd immunity, no epidemic*	0.5	3	0.75

Table 1 - A threshold game with risky prospects, illustrated in terms of a vaccination problem.

Before proceeding we note that, beside the vaccination problem, other risky interactions in everyday life present strategic features that to some extent are consistent with our toy game. For example, think of an individual's decision of whether or not to use protection during sex with an occasional partner. Alternatively, consider the case of a dissident deliberating on whether to engage in organized political activity. Let's suppose it is common knowledge that historically – conditional on the government being left unchecked – the executive may or may not take an authoritarian turn with equal probability. From the viewpoint of a citizen, participation in political activity is obviously costly; so, one would rather have other people take action than do it oneself. On the other hand, if one expects much of the population to remain inactive, one might be inclined to eventually engage in political action so as to avoid the gloomy prospect of an authoritarian regime. In all such cases, individuals' best-responses and hence the equilibria of the game will vary with the distribution of individuals' risk preferences (to be discussed in section III below).

2. *Experimental design and procedures*

Our experimental sessions were conducted at the University of Pennsylvania's Wharton Behavioral Lab. Upon arrival at the lab subjects were randomly allocated to computer terminals, where they expressed their consent to participate in an interactive decision-making experiment. On average, a session had about 17 subjects and lasted about 50 minutes. Each session consisted of the following stages: Introduction Stage; Play Stage; Payment Stage.

Below we describe the "main treatment".

Introduction Stage. After granting consent, subjects were asked to read the on-screen instructions; they were informed that they would go through a set of decision tasks, where each participant would be prompted to choose one of the actions represented by options on the screen, labeled as "A", "B", and "C". Each subject was instructed that any money she would earn (in addition to a flat participation fee) depended on her choice and on the choices made by all other participants in the lab session, as well as on the outcome of a fair coin tossed by the computer. In particular, each subject was informed that the choices made by all other participants – together with the outcome of a coin flip – would determine one of three scenarios. After reading the instructions, subjects were prompted to answer a set of comprehension questions.

Before moving on to the Play Stage, a few comments are due. First, we stress that participants' actions were simply denoted by "A", "B", "C" (i.e., in describing the scenarios, no reference was made to vaccination, free riding, etc.). Second, *letter-outcome pairs* (e.g., whether B is associated with the socially-undesirable option rather than, say, the exit option) were *randomized across participants*. This was done in order to control for the fact that letters that come first in the alphabet may be perceived as more prominent. For an instance of the experimental instructions featuring alternative letter-outcome pairs, please refer to the *Appendix*.

Play Stage. All plays were conducted using Behavery (https://behavery.com/): a software for laboratory, online, and field experiments. The order of subsequent tasks was as reported below.

(i) Each subject was asked to choose one of the options "A", "B", or "C". Subjects were instructed that, after all participants had made their choices, a fair coin would be tossed by the computer and the scenario for the current play would be determined (i.e., the same scenario for all participants). Note that subjects were not informed of the scenario they were in, either before or after making decisions.

(ii) Each subject was prompted to guess how many participants in the same session chose the option corresponding to the socially-undesirable action. Thus, (in the case of the letter-outcome pairs of Table 1 above) the task read as follows: *"... indicate the percentage of the participants in the entire room that you believe have chosen B..."*. Subjects entered their guesses by positioning a slider to the desired percentage. Upon doing so, they were informed that they would receive an additional payment of $0.25, if

they provided an accurate estimate within ± 1 percentage point of the realized value (and would receive nothing otherwise).[8]

(iii) "Part 2" instructions: subjects were told that they would go through (an unspecified number of) additional rounds involving the *same* decision task; in each round, the scenario would be determined by the new round's population-level behavior and new coin flip. (Subjects were told that only at the end of the experiment they would learn about the money earned over the rounds.) Further, at this point subjects were informed about the networked structure of the population, in the following way. First, subjects were told that each participant in the room was connected to some others at random, such that everyone was either directly or indirectly connected to everyone else (participants directly connected to one another were referred to as *neighbors*). Second, subjects were *not* informed about the specific number of connections they had, but they were just told that their neighbors were the *same* across rounds; additionally, subjects were informed that their neighbors might or might not have the same number of connections as they did. (In fact, unbeknown to subjects, the experiment's software was coded to randomly generate a network for each lab session, such that each node in a session had a degree centrality of either 2 or 3; this design feature guarantees sufficient variability in the feedback passed on to subjects, while ensuring that subjects are fully comparable across sessions or positions. See Figure 1 below for a sample network.) The above exhaustively outlines what subjects knew about the network.

(iv) Before carrying out the choice task in round 2, each subject was given feedback about the percentage of her neighbors that chose the socially-undesirable action in round 1; e.g., *"0.0% of your neighbors chose B in the previous round"*.

…

(v) Round t (**choice** task): each subject was asked to choose an option ("A", "B", or "C").

[8] The slider was initially positioned at a value of 50%. Subjects could not leave the slider in the initial position; so, they had to take a stance and express their beliefs about the frequency of the one action in relation to which the threshold is defined (i.e., the socially-undesirable action). For discussion on the merits of incentivizing the elicitation of beliefs, see Trautmann and van de Kuilen (2015): in particular, we stress that it has been shown that relatively small incentives for beliefs do not typically create a meaningful hedging opportunity (note that in our case the bracket for an accurate guess is 1 percentage point in either direction of the realized value).

(vi) Round *t* (**belief** elicitation): each subject was prompted to guess the percentage of participants in the *entire room* that she believed chose the socially-undesirable action in the current round *t*.

(vii) Round *t*+1 (**feedback** re. round *t*): each subject was given feedback about the percentage of her *neighbors* that chose the socially-undesirable action.

(viii) Round *t*+1 (**choice** task): each subject was asked to choose an option ("A", "B", or "C").

(ix) Steps *vi.* to *viii.* were repeated a number of times; i.e., subjects played 10 rounds in total.

(x) Subjects were given a brief demographic questionnaire. (For a copy of the experimental instructions, please see the *Appendix*.)

Payment Stage. The payment mechanism consisted of two parts: each subject received a flat $10 participation fee, in addition to any payoffs earned over the ten rounds (if positive). Each subject was informed that if the sum of her payoffs earned across rounds was negative, then she would only receive her $10 participation fee. (Whereas this payment system might in principle encourage risky choices, as will be clear it does not at all affect our analysis, which revolves around treatment effects.)

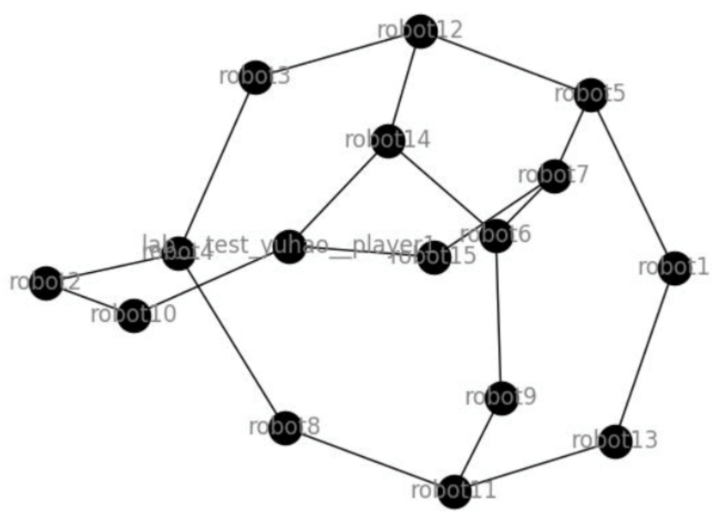

Figure 1 - A random network generated by Behavery (https://behavery.com/) as a simulation of the lab environment. Note: experimental subjects were *not* informed about their position in the network or about the specific network structure. Subjects were simply told that their neighbors were the same across rounds, and that their neighbors might or might not have the same number of connections as they did.

Finally, a "control treatment" was designed to further delve into individuals' risk-taking behavior in the domain of strategic interaction. The control treatment is the same as our main treatment, except that participants receive *no feedback* about their neighbors' choices; note that this is a between-subjects design. (We ran 6 sessions of the main treatment, and 5 sessions of the control treatment; no subject was allowed to participate in more than one session.)

III. Theoretical predictions and experimental hypotheses

We begin by considering equilibrium behavior in our control treatment. Since at the end of each round subjects received no information (in regards to their payoffs or otherwise), a session may be viewed as a series of one-shot games. For ease of reference, in what follows we refer to the actions in accordance with the letter-outcome pairs of Table 1 above. In particular, actions B and C respectively denote the *socially-undesirable* ("free ride") and *riskless* ("exit") options.[9]

Now, in order to formulate predictions in the context of our control treatment, one has to make some assumptions about the distribution of risk preferences in the population of participants. So, we shall first assume that all subjects are *risk-averse*, with an individual being said to be averse to risk "if she prefers the sure prospect with value *m* over any risky prospect with expected value *m*". (We later examine cases in which subjects are not necessarily averse to risk.) For simplicity, we assume preferences to be common knowledge. These assumptions imply the following claim.

Observation 1. *If all individuals are risk-averse, then all Nash equilibria in pure actions consist of the following profile:* $\bar{n}$ *players choose B ("free ride"), whereas* $n - \bar{n}$ *choose C ("exit"), with* n *denoting the total number of players in the population and* $\bar{n}$ *being equal to the largest number of players that corresponds to a fraction of the population < 40%.*

Proof. Let's consider two cases. [Case I] Suppose that current population-level behavior is described by a profile of actions whereby less than 40% of the population plays B. In that case, a risk-averse individual who currently plays C would rather deviate to B, as long as her deviation does not contribute to the population reaching the 40% threshold. By contrast, a risk-averse

[9] Action B of Table 1 is socially-undesirable to the extent that playing B induces a Pareto-dominated outcome (i.e., when 40% or more of the population plays B, as in the third class of equilibria we present below). Still, for ease of reference, we generically refer to that option as the "socially-undesirable action".

individual who currently plays B would have no incentive to deviate. [Case II] Alternatively, suppose that current population-level behavior is described by a profile of actions whereby 40% or more of the population plays B. In this case, a risk-averse individual would rather deviate to C (in fact, conditional on the others' choices, B and C have the same expected value, but C is the sure prospect). As a result of this deviation, either Case I or II will follow. So, ultimately in equilibrium $n - \bar{n}$ players choose C while $\bar{n}$ players choose B.

Next, if one assumes that some (or all) individuals may not be risk-averse, then several classes of equilibria exist. Thus, each of the following observations differs from the others in the fraction of the population that is risk-averse; moreover, observation 3 qualifies risk attitudes in terms of loss aversion.[10] (Note: since at $\boldsymbol{\omega_1}$ risk-neutral individuals are perfectly indifferent between the three actions – see p. 7 above – without loss of generality below we shall focus on cases in which individuals behave as either risk-averse or risk-seeking players.)

Observation 2. *If less than 40% of the population is risk-seeking and the rest is risk-averse, then $\bar{n}$ players choose B ("free ride"), whereas $n - \bar{n}$ choose C ("exit"); in particular, all the risk-seeking individuals play B, and the rest of the population plays B or C. (Note that, like before, n denotes the total number of players in the population and $\bar{n}$ is equal to the largest number of players that corresponds to a fraction of the population < 40%.)*

Proof. Let's consider two cases. [Case I] Suppose that current population-level behavior is described by a profile of actions whereby less than 40% of the population plays B. In that case, a risk-averse individual who currently plays C would rather deviate to B, as long as her deviation does not contribute to the population reaching the 40% threshold. By contrast, a risk-seeking individual would deviate to B regardless. [Case II] Alternatively, suppose that current population-level behavior is described by a profile of actions whereby 40% or more of the population plays B. In this case, a risk-averse individual would rather deviate from B to C. As a result of this deviation, either Case I or II will follow. So, ultimately in equilibrium $n - \bar{n}$ players choose C while $\bar{n}$ players choose B, with all the risk-seeking individuals playing B.

Observation 3. *If 40% or more of the population is risk-seeking, then any action profile whereby more than $\bar{n}$ players choose B ("free ride") and the rest A ("contribute") or C ("exit") may be an equilibrium, depending on the individuals' specific attitudes toward risk and loss.*

[10] For accounts of loss aversion, see among others: Kahneman and Tversky (1979), Köbberling and Wakker (2005), Kőszegi and Rabin (2006).

(Like before, n denotes the total number of players in the population and n̄ is equal to the largest number of players that corresponds to a fraction of the population < 40%.)

Proof. The proof is similar to the previous ones, and therefore is only sketched. We simply note that if 40% or more of the population plays B, then a loss-averse individual with reference point $0.75 (i.e., the payoff from the sure prospect C) will prefer C to A, and B to C, and hence choose action B. On the other hand, a loss-averse individual with reference point $0 (i.e., the payoff prior to playing the game) may well prefer A to C, and C to B, and hence choose action A.

Having analyzed equilibrium behavior in the control treatment, we turn to discuss the predicted impact of our main treatment on subjects' choice behavior. (This treatment manipulation will allow us to check if, under minimal assumptions, we can rule out failures in belief updating as an explanation for violations of expected utility theory in our games.[11]) To that end, we define the "*low-feedback sample*" as the group of subjects who – in a given round – receive feedback (about the frequency of B choices in their neighborhood) below the threshold, i.e., < 40%. Similarly, we define the "*high-feedback sample*" as the group of subjects who, in a given round, receive feedback above the threshold.[12] Given this, we make the assumption that participants update their beliefs about the population-level frequency of B choices, on the basis of the feedback (about the neighborhood-level frequency of B choices). Regardless of the specific distribution of individual-level (risk or ambiguity) attitudes in the population, but provided it is the same across samples, Bayesian rationality implies the following prediction.

H1: participants in the *high-feedback sample* choose the riskless action (C) weakly more frequently than participants in the *low-feedback sample*.

To grasp the basic intuition behind H1, we begin by making an informal argument (which we formalize below). For the time being, consider a participant in the main treatment who – before

[11] We stress that our experiment is *not* designed to analyze learning in relation to structural properties of a network; as such, our design does not aim to disentangle Bayesian and "naïve" belief-updating processes (DeGroot, 1974). For recent network-theoretic accounts of learning, we refer the reader to: Gale and Kariv (2003); Golub and Jackson (2010); Acemoglu, Dahleh, Lobel, and Ozdaglar (2011); for a survey, see Jackson and Yariv (2011). We further note that models of social learning typically focus on equilibria with risk-neutral players. Thus, our exercise rests on broadly different assumptions than those made by experimental tests of social learning (see among others Chandrasekhar, Larreguy, and Xandri, 2020, and Grimm and Mengel, 2020).

[12] We take the opportunity to stress that "feedback" refers to the information about *neighbors*, not the entire population of players (participants in the room). On the other hand, payoffs depend on whether the frequency of the socially-undesirable action exceeds the threshold at the *population* level, not simply at the neighborhood level.

receiving feedback – states a belief indicating that the proportion of B choices was below the 40% threshold. We note that such a subjective probability entails that the subject believes that *no risky scenario* will likely occur: this is because risky scenarios (Y or Z) occur if and only if 40% or more of all participants choose action B. Now, suppose that the subject subsequently receives high feedback. Naturally, her posterior beliefs will go up; in fact, depending on the priors, posteriors might even rise above the threshold, *which would entail a risky scenario* ($\boldsymbol{\omega_1}$). In that case, unless one is a risk-seeker, following the high feedback one would rather choose the riskless action (C). Hence, on average we expect participants in the high-feedback sample to collectively choose the riskless action weakly more frequently than would participants in the low-feedback sample.

More formally, note that from a subject's viewpoint the *feedback* follows a binomial distribution $Bin(\eta, \vartheta)$, where η denotes the sample size (corresponding to the subject's degree centrality, i.e., the number of randomly-assigned neighbors) and ϑ denotes the rate of B choices in the entire population. Bayesian inference involves modeling unknown parameters as random variables. So, one can model ϑ as a random variable, where a *prior* distribution represents the subject's initial beliefs about the possible values of the parameter ϑ.[13] One can then use the Bernoulli distribution to obtain the *likelihood* of the feedback received, for all possible values of ϑ (i.e., use a "Bernoulli likelihood function"). Finally, Bayes' theorem implies that $p(\vartheta \mid feedback) \propto p(\vartheta) \cdot p(feedback \mid \vartheta)$: that is, the posterior $p(\vartheta \mid feedback)$ is proportional to the prior $p(\vartheta)$ times the likelihood $p(feedback \mid \vartheta)$; DeGroot and Schervish (2012, pp. 390-391).[14] Intuitively, this implies that the higher is the feedback, the higher is the posterior.

Given the above, and assuming that the average distribution of priors is the same across low/high feedback samples, it follows that participants in the *high-feedback sample* should hold higher posterior beliefs (about the rate of B choices) than those in the *low-feedback sample*. This means that participants in the high-feedback sample are more likely to hold posteriors exceeding the threshold, thereby entailing a risky scenario. Hence, under the assumption that the number of

[13] Note that the experimental design elicits a simple guess, that is, the percentage of participants that the subject believes have chosen the socially-undesirable action (*"... indicate the percentage of the participants in the entire room that you believe have chosen B…"*). Now, suppose that in response to that a subject enters a value of, say, 33%. That may be interpreted in two different ways: (i) as a statement that the subject *believes with certainty* that one third of the participants chose B; (ii) as a statement that the subject *believes it most likely* that one third of the participants chose B. The latter interpretation appears to be more realistic, and justifies modeling the prior as a distribution rather than a point estimate.

[14] See also Gilboa and Schmeidler (1993) for an axiomatization of Bayesian updating, given ambiguous beliefs.

risk-averse subjects is similar across samples, it follows that participants in the high-feedback sample should choose the riskless action (C) weakly more frequently than participants in the low-feedback sample. As we shall see, this prediction is supported by our behavioral data.

Our next hypothesis delves into subjects' beliefs. As we discussed, Bayesian rationality implies that the higher is the feedback, the higher is the posterior. This prediction may be tested directly, by examining our *stated beliefs* data.

H2: participants' stated beliefs are positively related to the magnitude of the feedback received.

As will be shown, the above hypothesis is supported by the data, in that our results indicate an impact of the preceding round's feedback type (i.e., low/high sample) on the belief-formation process, on average.

Our final hypothesis below focuses on participants stating a belief that indicates that the proportion of B choices was above the 40% threshold: below we shall refer to such subjects as high-belief (*"HB"*) participants. Note that stating an above-threshold subjective probability entails that *one believes that a risky scenario (Y or Z) will likely occur*. Now, suppose that an HB-participant subsequently receives high feedback. As a result of the high feedback, that subject might update her beliefs upwards. However, such a belief update should not entail a different scenario relative to the previously-stated belief. This is because risky scenarios (Y or Z) occur if and only if 40% or more of the population chooses action B. Since – in this case – both priors and posteriors exceed the threshold, they ultimately imply the same scenarios. This leads to the following prediction.

H3: HB-participants in the *high-feedback* sample choose the riskless action (C) with the same frequency as HB-participants in the *control* (i.e., no feedback treatment).

As we shall see, this prediction is falsified by the data, as HB-participants in the high-feedback sample happen to choose the riskless action significantly more frequently than HB-participants in the control. We attribute this apparent preference reversal to people's limitations in performing hypothetical reasoning. (As the feedback slightly attenuates one of the sources of complexity of the problem – i.e., strategic uncertainty – some subjects appear to focus on the other source of complexity, namely, risky prospects.)

IV. Experimental results

1. *Main treatment: summary statistics, preliminary tests*

We begin by addressing our main treatment (comprising the low- and high-feedback samples). A total of 101 subjects from several academic departments took part in our sessions at the University of Pennsylvania's Wharton Behavioral Lab; the mean age was 24.7 years. Subjects on average earned a total payoff of $11.31 (over ten rounds), in addition to the $10 participation fee. On average subjects chose the option associated with actions A, B, and C of Table 1 above 10.59%, 56.83%, and 32.58% of the time, respectively. Further, a minority of subjects chose the same action in every round; i.e., out of a total of 101 participants in the main treatment, about 1%, 27%, and 10% of the subjects respectively chose A, B, and C in each and every round.

The first two columns of Table 2 report average beliefs (about the *population-level* frequency of B choices) held by the subjects who in a given round chose the option indicated on the left-hand side of the table. The last column of Table 2 reports the mean feedback (about the *neighborhood-level* frequency of B choices) provided to subjects who chose each of the options on the left-hand side. Notably, in all cases, beliefs and feedback are above the 40% threshold on average.

	Beliefs about population: *Round 1*	Beliefs about population: *Other rounds*	Feedback about neighbors: *Other rounds*
A	54.36	59.04	60.95
	(27.51)	(23.26)	(33.37)
B	59.47	60.62	54.62
	(21.641)	(23.16)	(30.66)
C	56.20	60.23	62.70
	(20.40)	(21.24)	(28.39)

Table 2 - Main treatment. Mean beliefs and feedback (about B choices) held by the subjects who in a given round chose the option indicated on the left-hand side of the table; in parentheses is the standard deviation. Note: 101 subjects took part in the main treatment. We report the beliefs held in round 1 in a separate column, as subjects stated such beliefs before receiving any feedback; naturally, no feedback about previous play was provided in round 1.

For the purpose of providing a more granular depiction of the data, Figure 2 below breaks down the (mean) values of each of the following variables, by round: (i) frequency of B choices; (ii) beliefs about population-level B choices; (iii) feedback about neighborhood-level B choices.

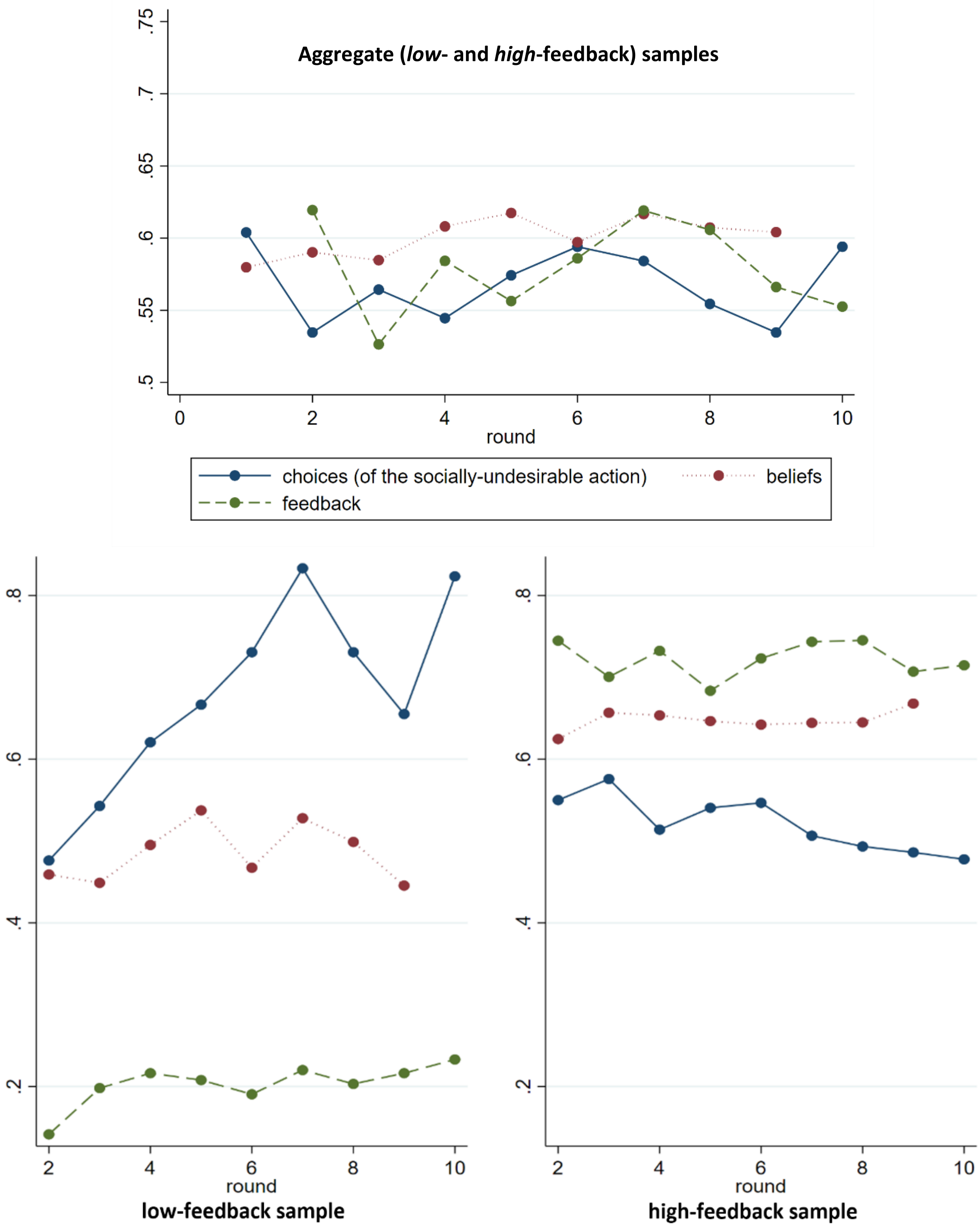

Figure 2 - Main treatment. The upper panel shows line graphs depicting mean values (by round) of: the frequency of B choices (i.e., free riding), beliefs about population-level B choices, feedback about neighborhood-level B choices. The lower panel breaks down *low*- vs *high*-feedback samples. Note: no feedback about previous play was provided in round 1; no beliefs were elicited after the last choice task was carried out (in round 10). For the sequence of experimental tasks, see p. 11.

Now, in order to give a rough overview of some suggestive data patterns, we first report non-parametric tests and later assess our hypotheses formally (by means of a regression analysis). For the time being, as a very preliminary test of the equality of choice distributions across low- versus high-feedback samples, we verify some hypotheses via pairwise comparisons. In short, a test of proportions (adjusted for clustering on 101 subjects, using data from all the rounds in which feedback was provided; i.e., rounds 2-10) indicates that the socially-undesirable action B was chosen more often in the *low*- than in the *high*-feedback sample; that is, respectively 67.73% and 52.13% of the time ($z = 2.25$, $p = 0.024$, two-tailed). In other words, subjects were *less likely* to choose B after learning that the *neighbors chose B in proportions larger than the threshold*.

Furthermore, the same test shows no meaningful differences in the proportions of choice of action A across samples, which were respectively 10.36% and 10.64%. By contrast, performing the same test with respect to action C reveals that the riskless option was chosen less often in the *low*- than in the *high*-feedback sample; that is, respectively 21.91% and 37.23% of the time ($z = -2.25$, $p = 0.024$, two-tailed). Taken together, these data patterns provide some preliminary evidence that exposing subjects to high feedback might cause them to shift from a risky action (B) to a riskless action (C).

2. *Main treatment: regression analysis—effect of feedback on risk-taking behavior*

The tests above capture across-rounds average trends, but do not account for differences in the beliefs held by subjects or other differences across rounds (such as differences in the values of the feedback itself). For those reasons, we now turn to the regression analysis.

We start by examining one's choice of the riskless action C (against its negation, i.e., "not C"). Specifically, the models in Table 3 below present the results of logit regressions consisting of one's choice of the riskless action as the binary dependent variable, and some or all of the following predictors: a low/high feedback indicator; dummy variables indicating whether one's stated belief is below/above the threshold at round 1, or at round t; and finally a time (i.e., round t) variable. We note that the standard errors are adjusted for clustering on 101 subjects.

Model 1 in Table 3 below confirms that the low/high feedback indicator significantly affects one's choice of the riskless action C. This formally provides evidence in support of H1, which states that *participants in the high-feedback sample should choose the riskless option weakly more frequently than participants in the low-feedback sample*.

choice of action C at *t*	[1]	[2]
low/high feedback indicator (*t*-1)	.748*** (.269)	.537** (.267)
belief dummy (*t*)		.226 (.270)
belief dummy (at round 1)		.196 (.414)
round (*t*)		.057** (.025)
constant	-1.270*** (.263)	-1.776*** (.402)
Pseudo R2	0.017	0.017
AIC	1136.759	1017.101
Obs.	909	808

Table 3 - Logit coefficients of two models estimating one's choice of the riskless action (C) at round *t* of the main treatment. In parentheses are robust standard errors clustered on 101 subjects (*, **, and *** respectively indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, for the relevant Z-statistic, two-tailed tests). Model 1 uses all choice tasks except for round 1, for which there was no t-1 feedback. Model 2 uses the same data as model 1 except for round 10, as no beliefs were elicited after the last choice task; also note that this model includes the belief held in round 1 as a separate predictor, since subjects stated such beliefs before any feedback (hence, may be viewed as priors).

As a robustness check, model 2 in Table 3 includes both beliefs and time variables, as predictors. (Note that model 2 includes the belief held in round 1 as a separate predictor, since subjects stated such beliefs before receiving any feedback and so they may be viewed as priors.) Model 2 confirms the significant impact of the low/high feedback indicator; moreover, it shows no significant effects for the other predictors, except for the round variable, which may reflect differences in the feedback passed on to subjects across rounds. In summary, the regressions fully support H1: when subjects learn that neighbors chose B in proportions larger (smaller) than the threshold, subjects are more (less) likely to choose the riskless action C in the next round.

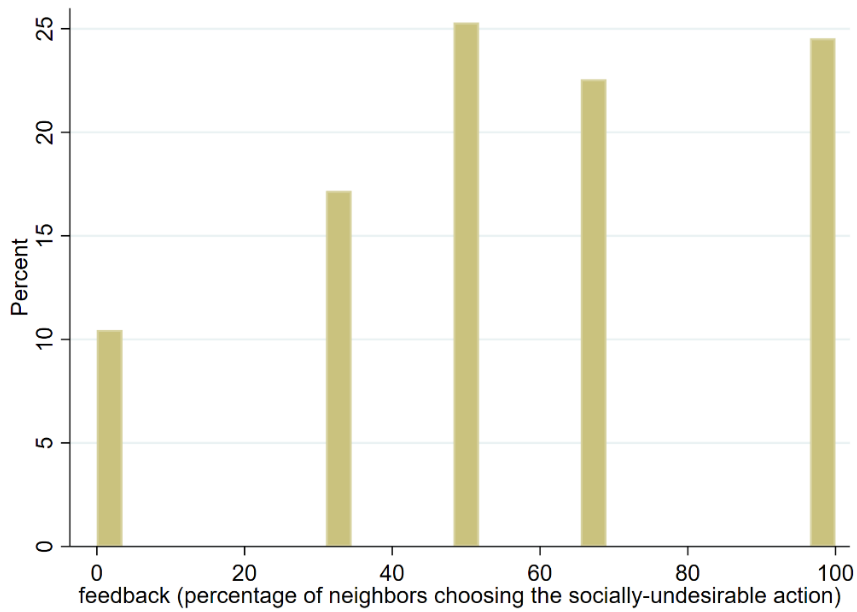

Figure 3 - Histograms for the feedback that was passed on to subjects. As an example, note that the first bar indicates that about 10 percent of the time subjects were informed that none (0%) of their neighbors chose the socially-undesirable action B in the preceding round. The relative frequency of each feedback value is calculated with respect to all the rounds in which feedback was provided, namely, rounds 2-10. Note: by design the feedback could take on five possible values, since the experiment's software was coded to randomly generate a network per session, such that each node had a degree centrality of either 2 or 3. (Subjects were not informed about the specific network structure; however, they were informed that their neighbors were the same across rounds; for details, see pp. 10-11 above.)

We move on to an additional robustness check. We note that the estimates above may be biased by individual-specific characteristics, such as one's risk and ambiguity attitudes; also, the estimates might vary with the actual feedback values one receives across rounds. Hence, for robustness purposes here we consider a conditional (fixed-effects) logit model.

Before presenting the model, we note that – by design – conditional logit models cannot use observations that have no variation, such as panel units (i.e., subjects) making the same choice in each and every round; such observations are therefore automatically dropped from the regression, thereby resulting in a smaller sample.[15] This explains why this model is presented as a robustness check, rather than the main analysis.

[15] Specifically, (since the dependent variable is dichotomous) subjects who chose the riskless action *in every round* are dropped, as are subjects who chose the riskless action *in no round at all*.

choice of action C at *t*	
low/high feedback indicator (*t*-1)	1.706** (.680)
feedback value (*t*-1)	.007 (.016)
high feedback indicator * feedback value	-.020 (.017)
Pseudo R2	0.022
Obs.	441

Table 4 - Conditional (fixed-effects) logit coefficients of a model estimating one's choice of the riskless action (C) at round t of the main treatment. In parentheses are standard errors (*, **, and *** respectively indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, for the relevant Z-statistic, two-tailed tests). For a given subject, the model uses all choice tasks except for round $t = 1$ (for which there was no feedback). By design, panel units (i.e., subjects) making the same choice in each and every round are automatically dropped. Also note that conditional logit models have no constant term.

That said, Table 4 presents a fixed-effects logit model consisting of one's choice of the riskless action as the binary dependent variable, and of the following predictors: (*i.*) a low/high feedback indicator; (*ii.*) the numerical feedback variable (i.e., the actual feedback values passed on to subjects, as opposed to the low/high feedback indicator); (*iii.*) the interaction of *i.* and *ii.*

In short, Table 4 corroborates our previous evidence, showing that the low/high feedback indicator significantly affects one's choice of the riskless action C. On the other hand, the actual feedback value one is exposed to does not appear to significantly affect one's choice, nor does the interaction variable.[16] This provides the ultimate evidence in support of H1: when subjects learn that neighbors chose B in proportions larger (smaller) than the threshold, subjects are more (less) likely to choose the riskless action C in the next round. In concluding, we stress that the estimated coefficients of fixed-effects models *cannot* be biased because of omitted time-invariant characteristics, such as individual-level risk or ambiguity attitudes.

[16] Incidentally, note that adding a time (i.e., round t) variable as a predictor does not qualitatively affect the results; in fact, the time variable turns out to be non-significant when added to the predictors in Table 4.

3. *Main treatment: effect of feedback on stated beliefs*

Bayesian rationality entails that *high* feedback is more likely to induce a higher posterior than *low* feedback, all else equal. This prediction was indirectly confirmed by the results above. Yet, it may be tested directly by examining our *stated beliefs* data. So, we now turn to analyze subjects' beliefs at round t. To that end, we present a fixed-effects regression consisting of one's stated belief as the continuous dependent variable, and some or all of the predictors as the model above, that is: (*i.*) a low/high feedback indicator; (*ii.*) the numerical feedback variable (i.e., the actual feedback values passed on to subjects); (*iii.*) the interaction of *i.* and *ii.*

stated belief at t	[1]	[2]
low/high feedback indicator (t-1)	6.278*** (1.709)	6.317* (3.493)
feedback value (t-1)		.139 (.086)
high feedback indicator * feedback value		-.099 (.092)
constant	55.732*** (1.373)	52.856*** (2.191)
R-squared	0.103	0.146
AIC	6688.83	6688.933
Obs.	808	808

Table 5 - Fixed-effects regression coefficients of models estimating one's stated belief at round t of the main treatment; standard errors are shown in parentheses (*, **, and *** respectively indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, for the relevant Z-statistic, two-tailed tests). The regressions are performed on the dataset containing all observations except for round $t = 1$ (as there was no t-1 feedback) and round $t = 10$ (since no beliefs were elicited after the last choice task).

Model 1 in Table 5 reveals that the low/high feedback indicator positively and significantly affects one's belief-formation process at t. Put differently, when subjects learned that neighbors chose B in proportions larger than the threshold, subjects were likely to state higher beliefs about the (population-level) frequency of B choices. This provides evidence in support of H2, which says that *participants' stated beliefs are positively related to the magnitude of the feedback received*.

As a robustness check, model 2 includes the numerical feedback variable and the (feedback *indicator*value*) interaction variable. In brief, model 2 confirms a (mildly) significant impact of the preceding round's feedback type (i.e., low/high) on the belief-formation process, corroborating the previous evidence. Incidentally, we note that when controlling for the feedback type, the actual feedback values passed on to subjects do not appear to significantly affect beliefs (this might indicate that different feedback values are generally interpreted differently by different subjects). Overall, the models show that subjects do use the feedback to inform beliefs; later we shall further delve into our subjects' beliefs and their impact on behavior.

4. *Control treatment: summary statistics and non-parametric tests*

We proceed to discuss results from our control treatment. This was identical to the main treatment above, except that participants received *no feedback* about their neighbors' choices. A total of 84 subjects took part in the control treatment; the mean age was 23.7 years, and other demographics were similar across treatments. Participants in the control on average earned a total of \$6.61 (over ten rounds), in addition to the \$10 participation fee.

Note that the payoff data signal some sharp differences across treatments: as the reader might recall, participants in the main treatment made about twice the money as in the control (i.e., \$11.31, plus the participation fee). A two-tailed Wilcoxon-Mann-Whitney test conducted on the entire subjects sample confirms that payoffs were significantly lower in the control than in the main treatment (N = 185 subjects, Z = -5.275, p = 0.000). In what follows we investigate what drove such a difference in earnings.

We begin by reporting summary statistics relating to the choice data. On average, participants in the control chose the option associated with actions A, B, and C of Table 1 above 9.64%, 64.29%, and 26.07% of the time, respectively. (In the main treatment, the same actions were chosen respectively 10.59%, 56.83%, and 32.58% of the time.) Also, a non-trivial fraction of subjects chose the same action in each and every round: that is, out of a total of 84 participants in the control treatment, about 0%, 36%, and 7% of the subjects respectively chose A, B, and C in every round; these figures appear somewhat different than the corresponding figures from the main treatment (1%, 27%, and 10%, respectively). Below we shall formally contrast data from our two treatments.

	Beliefs about population: *Round 1*	Beliefs about population: *Other rounds*
A	56.33	52.03
	(18.40)	(21.50)
B	65.51	64.66
	(20.49)	(22.32)
C	55.04	49.12
	(20.83)	(18.15)

Table 6 - Control treatment. Mean beliefs (about B choices) held by the subjects who in a given round chose the option indicated on the left-hand side of the table; in parentheses is the standard deviation. Note: 84 subjects took part in the control treatment. To be consistent with the style of Table 2 above, we report the beliefs for round 1 in a separate column.

Table 6 reports average beliefs (about the *population-level* frequency of B choices) held by the subjects who chose the option indicated on the left-hand side of the table. See also Figure 4 below, which shows line graphs of choices and beliefs (versus round). In short, average beliefs held by participants in the control treatment are well above the 40% threshold, as was in fact the case in the main treatment.

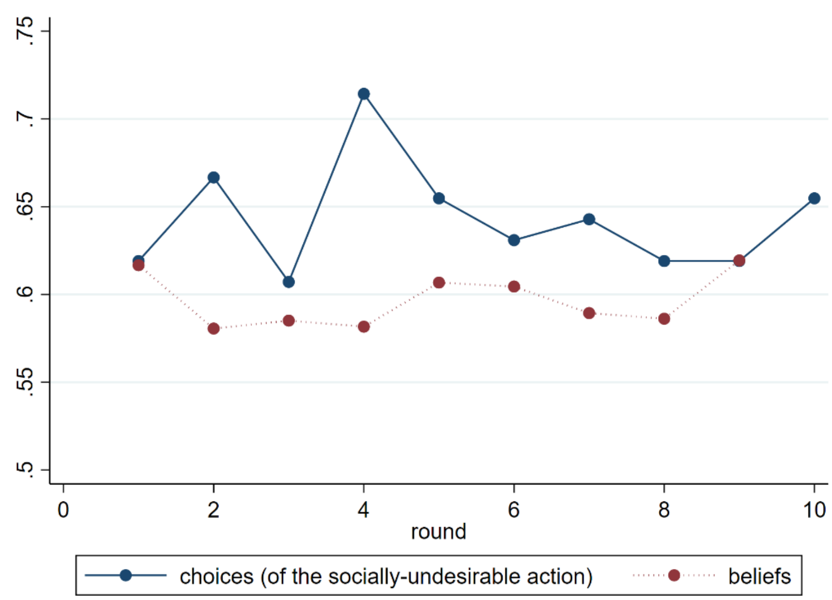

Figure 4 - Control treatment. Line graphs depicting mean values (by round) of: the frequency of B choices (i.e., free riding) and beliefs about population-level B choices.

Yet, comparing Figure 4 with (the upper panel of) Figure 3 reveals that even though average beliefs appear similar between treatments, participants in the control treatment chose B in proportions greater than in the main treatment. In fact, the upper panel of Figure 3 suggests that exposing subjects to feedback *in aggregate* causes them to shift away from a risky action. Thus, moving from the main treatment to the control, we might expect an increase in risky (B) choices along with a fall in riskless (C) choices: this informs the next directional hypotheses.

A one-tailed test of proportions clustered on (185) subjects indicates that the risky action B was chosen mildly-significantly *more often in the control* than in the main treatment (z = 1.463, p = 0.071; the test is conducted on the entire sample of observations except round 1, for which there was no feedback in the main treatment). The same test shows that the riskless action C was chosen mildly-significantly *less often in the control* than in the main treatment (z = -1.339, p = 0.090). In summary, these data patterns provide some rough, preliminary evidence that participants in the control treatment behaved *as if* they were more risk-seeking than participants in the main treatment. The analysis below will shed light on these findings.

5. *What is rational and what isn't*

In what follows, we shall formally verify if differences in risk-taking behavior between treatments may or may not be ascribed to differences in beliefs across treatments (with and without feedback). In other words, we investigate whether any differences in behavior may be rationalized as the *effect of a belief revision on an individual's expected utility*. To that end, the model in Table 7 below presents the results of a logit regression consisting of one's choice of the riskless action as the binary dependent variable, and of the following predictors:
(*i.*) a control/main treatment indicator; (*ii.*) a dummy variable indicating whether one's stated belief is below/above the threshold; (*iii.*) the interaction of *i.* and *ii.* As in our non-parametric tests above, the standard errors are adjusted for clustering on 185 participants (i.e., 101 in the main treatment, plus 84 in the control treatment).

In brief, the model in Table 7 reveals no significant differences between treatments; by contrast, it reveals a significant impact of both the *belief dummy* and the (*treatment***belief*) interaction variable. Note that the significance of the belief dummy means that – regardless of the treatment – one's choice of the riskless action C depends on whether one believes that the proportion of B choices is above the 40% threshold. Moreover, conditionally on stating such a belief, one's choice of the riskless action is more likely if one is assigned to the main treatment (the significance of the interaction variable deserves further investigation, to which we shall attend later).

It is worth noting one more point about the *belief dummy* (Table 7): importantly, its significance confirms the presence of subjects with (standard-selfish) rational preferences in our dataset. To see why, note that such preferences rule out the so-called "risk contagion" effect, whereby subjects are more prone to choose a risky option *if* they believe others are making risky choices (for evidence of this effect, see among others Reiter, Suzuki, O'Doherty, Li, and Eppinger, 2019).[17] In our game, rational preferences imply that action C is a best-response to a belief above the 40% threshold, unless one is a risk-seeker. So, the significance of the belief dummy (Table 7) confirms the presence of rational participants with risk-averse preferences in our data.

choice of action C at *t*	
control/main treatment indicator	-.551 (.369)
belief dummy	-.674** (.289)
main treatment indicator * high belief	1.101*** (.382)
constant	-.530* (.282)
Pseudo R2	0.012
Obs.	1,665

Table 7 - Logit coefficients of a model estimating one's choice of the riskless action (C) at round *t*. In parentheses are robust standard errors clustered on 185 subjects (*, **, and *** respectively indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, for the relevant Z-statistic, two-tailed tests). The regression is performed on the dataset containing all observations from the main and control treatments except for round $t = 10$ (since no beliefs were elicited after the last choice task).

[17] Specifically, contrary to our data, Reiter et al. (2019) show evidence of a "risk contagion" effect where subjects are more likely to choose a risky option (in the form of a lottery) over a sure option, if they observe others' risky choices. Similar results are reported by Gioia (2017) and by Fatas, Hargreaves Heap, and Rojo Arjona (2018), among others. A possible explanation for the risk contagion effect might be "conformist preferences", whereby one likes to do what one believes others will do (Charness, Naef, and Sontuoso, 2019; Goeree and Yariv, 2015).

We turn to discuss our final, key result. In order to verify whether differences in behavior may be – entirely – rationalized as the *effect of a belief revision* on an individual's expected utility, we move on to isolate observations for which a rational belief update should *not* imply a change in behavior. To do so, we consider participants (in either treatment) stating a belief that indicates that the proportion of B choices is above the 40% threshold: we refer to such subjects as *high-belief ("HB") participants*.

Recall that stating an above-threshold probability entails that one believes that a risky scenario (Y or Z) will likely occur. Thus, suppose that after stating her belief for the round, an HB-participant receives high feedback (in the main treatment). As a result of the high feedback, that subject might update her beliefs upwards. Yet, such a belief update should *not* entail a different scenario, relative to the previously-stated belief: since both priors and posteriors exceed the threshold, they ultimately imply the same scenarios![18] Under the assumption that individuals' characteristics – such as risk and ambiguity attitudes – are similar across treatments, our final prediction follows. Conditional on stating high beliefs, subjects who were provided high feedback (in the *main treatment*) should be as likely to choose C as subjects who were not provided any feedback at all (in the *control treatment*).

In order to test this prediction, the logit regression in Table 8 below is performed on the dataset containing observations relating to any round in which participants – in either treatment – stated a belief above the threshold (i.e., HB-participants). The logit model consists of the choice of the riskless action (C) as the binary dependent variable, and of the control/main treatment indicator as the sole predictor. Like before, the standard errors are adjusted for clustering on subjects.

Table 8 shows that participants in the main treatment take the riskless action more frequently than one would otherwise do in the absence of feedback, given high beliefs. This result provides evidence against H3, which states that *HB-participants in the high-feedback sample choose the riskless option with the same frequency as HB-participants in the control (i.e., no feedback treatment)*.

[18] Recall that risky scenarios (Y or Z) occur *if and only if* 40% or more of the population chooses action B; see p. 7.

choice of action C at *t*	
control/main treatment indicator	.625** (.268)
constant	-1.204*** (.194)
Pseudo R2	0.016
Obs.	1,212

Table 8 - Logit coefficients of a model estimating one's choice of the riskless action (C) at round *t*. The treatment indicator takes on value 1 when a subject belongs to the main treatment. In parentheses are robust standard errors clustered on 179 subjects (*, **, and *** respectively indicate $p < 0.10$, $p < 0.05$ and $p < 0.01$, for the relevant Z-statistic, two-tailed tests). Note: the regression is performed on the dataset containing all the observations relating to any round in which a subject – in either treatment – stated a belief above the threshold ("HB-participants"). 6 subjects do not meet this condition in any round.

Given the large sample size and the homogeneity of the demographic characteristics across treatments, we have no reason to assume any fundamental (prior) differences in individuals' characteristics – including risk and ambiguity attitudes – across treatments, on average. This suggests that our main treatment's manipulation (i.e., merely providing feedback about neighbors) seemingly causes a different appreciation of the available risky prospects.

We believe that some subjects' difficulty in processing risky prospects may be exacerbated by the strategic uncertainty that is intrinsic to a game with multiple equilibria. As the feedback lessens one of the sources of complexity of the problem (i.e., strategic uncertainty), some subjects seem to focus on the other source of complexity – namely risk – thereby exhibiting a different appreciation of the risky prospects. To sum up, we believe that the apparent "reversal" in risk preferences signals people's limitations in performing hypothetical reasoning in complex problems.

V. Concluding remarks

This paper has investigated "hypothetical reasoning" in problems where the impact of *risky prospects* (chance moves with commonly-known conditional probabilities) is compounded by *strategic uncertainty* (arising from the multiplicity of equilibria). Specifically, we consider "threshold games with risky prospects". Given these games, our design verifies if risk-taking behavior is affected by information that varies the extent of strategic uncertainty; to do so, our design exploits the fact that best-responses depend on individuals' beliefs, as well as on their risk preferences. This has permitted us to test some aggregate-level implications of expected utility theory.

Our results indicate a significant effect of the transmitted information on behavior: this effect is attributable in some cases to a *belief revision* about others' actions, and in other cases to a *reversal in risk preferences*. In order to isolate the latter, our analysis has identified observations for which a rational belief update (resulting from the feedback) should not imply a best-response other than the action one would choose in the absence of feedback. Comparing choice distributions across treatments, we have shown that participants in the main treatment take the riskless action more frequently than participants would do in the absence of feedback, given the same beliefs.

The above suggests that the experimental manipulation causes a different "understanding" of the problem's risky prospects. Unlike other violations of expected utility theory, this pattern is not rationalizable by models with non-standard (risk or ambiguity) attitudes. Thus, we impute this preference reversal to people's limited ability to perform hypothetical reasoning in complex tasks. As the experimenter's feedback slightly lessens one of the sources of complexity of the problem (i.e., strategic uncertainty), some subjects seem to focus on the other source of complexity (i.e., risky prospects).

Interestingly, recent research has documented *suboptimal* choices in non-strategic problems, as evidence of subjects' difficulties with hypothetical reasoning (Charness and Levin, 2009; Esponda and Vespa, 2014; Martínez-Marquina, Niederle, and Vespa, 2018) Similarly, in the context of experimental auctions, Levin, Peck, and Ivanov (2016) have shown that bidders often make mistakes, "due to a failure of insight or recognition, which would involve logical reasoning without requiring any explicit updating of probabilities" (Levin et al., 2016, p. 40). Our results complement previous findings by investigating a broadly different environment; in particular, our experiment suggests that difficulties with hypothetical reasoning may generally affect subjects' assessment of risk.

To conclude, this paper has studied a novel class of games (threshold games with risky prospects), and isolated a reversal in risk preferences in these games. In this connection, we note that several risky interactions in everyday life present strategic features that – to some extent – are consistent with our experimental environment. An application of particular interest to current policy-makers may be the case of an individual's choice of whether to get a vaccine on the verge of a possible epidemic. In that case, our study suggests that information campaigns alerting people of their neighbors' free-riding choices might discourage further risk-taking behavior. Although our toy game does not aim to capture the rich complexity of an actual vaccination decision, our study does provide intriguing insights in relation to a class of risky interactions, which may be worth investigating in the field.

APPENDIX

Experimental instructions and screen shots

NOTE: as we discuss in section II.2 above, *letter-outcome pairs* (e.g., whether B is associated with the socially-undesirable option rather than, say, the exit option) *were randomized across participants*. This was done in order to control for the fact that letters that come first in the alphabet may be perceived as more prominent. Below is an instance of the experimental instructions (for the main treatment) where the socially-undesirable option is associated with action A: accordingly, in the below screen shots, the threshold is defined in relation to action A; hence, the belief elicitation task and the feedback refer to action A. Finally, note that experimental instructions for the control treatment are the same as the main treatment, except that there is no feedback.

[Welcome screen]

At the beginning of this study, you will receive instructions on what to do and how your decisions can affect your earnings. Your participation in the study is voluntary. You may end your participation at any point, without loss of any benefits to which you are entitled.

The main purpose of the study is to explore people's decision making in different contexts. The study involves monetary decisions, that can only add to the $10 (show up fee) you receive for your participation. The duration of the study will be about 50 minutes.

Your final earnings depend on the decisions you and other participants make.

Please click the box if you agree to participate in the study.

Instructions (1/3)

You will receive a show up fee, and can earn additional money. The additional payment will be determined by your own choices and those made by the other participants, according to rules described below. Your final earnings will be added to your show up fee if positive.

In each round, each participant will be asked to choose one of the actions represented by options on the screen, namely "A", "B", and "C". Please note that the information about the amount of money earned over each round will be provided only at the end of the experiment.

Next

Instructions (2/3)

The money you will earn in each round depends on your choice, as well as on the choices made by all other participants, and on the outcome of a coin tossed by the computer in each round. The coin may result in either of two outcomes, HEADS or TAILS, each with a 50% chance. Depending on the conditions described above, you will end up in ONE of three alternative *scenarios*:

If less than 40% of all participants chose A, then *regardless of the coin outcome*:

- Your earnings for the round will be \$3.0 if you chose A, \$0.5 if you chose B, and \$0.75 if you chose C.

A	B	C
\$3.0	\$0.5	\$0.75

If 40% or more of all participants chose A, then:

- If the coin outcome is HEADS
  Your earnings for the round will be \$-1.5 if you chose A, \$1.0 if you chose B, and \$0.75 if you chose C.

A	B	C
\$-1.5	\$1.0	\$0.75

- If the coin outcome is TAILS
  Your earnings for the round will be \$3.0 if you chose A, \$0.5 if you chose B, and \$0.75 if you chose C.

A	B	C
\$3.0	\$0.5	\$0.75

Next

Instructions (3/3)

After all participants have made their choice, the coin is tossed by the computer, and the scenario for the round is determined.

(Participants will *not* be informed of the scenario they are in before making decisions.)

At any point during the experiment, if you have any questions please raise your hand and an experimenter will approach you.

Next

Control Questions

If more than 40% of all participants chose A, the coin outcome is HEADS, and you chose C, how much will you earn? 1.0

If less than 40% of all participants chose A, the coin outcome is TAILS, and you chose C, how much will you earn? -1.5

If less than 40% of all participants chose A, the coin outcome is TAILS, and you chose B, how much will you earn? -1.5

If more than 40% of all participants chose A, the coin outcome is HEADS, and you chose A, how much will you earn? -1.5

If less than 40% of all participants chose A, the coin outcome is TAILS, and you chose A, how much will you earn? -1.5

Hover (using mouse) and Scroll (using arrow keys) to review previous instructions

Next

You are currently in round 1 .

Choose an action from below
C A B

Hover (using mouse) and Scroll (using arrow keys) to review previous instructions

Next

Move the slider below to indicate the percentage of the participants in **the entire room** that you believe have chosen A in this round.

You will earn $0.25 if you guess within 2 percentage points (1 point in either direction) of the actual percentage.

%

0% 100%

Ending Round 1

Waiting for other participants...

Please Wait ...

Instructions part 2 (1/2)

In the following rounds you will face the same decision task as before.

Each participant in the room is connected to some others at random, such that everyone is either directly or indirectly connected to everyone else.

Participants who are directly connected to one another are “neighbors” (your neighbors are most likely not the participants sitting next to you).

Those who are indirectly connected to you are your neighbors' neighbors, the neighbors of your neighbors' neighbors, and so on.

Next

Instructions part 2 (2/2)

All connections (direct and indirect) remain constant across rounds. That is, if you are connected to specific participants in round 1, they will be your neighbors in all rounds.

Your neighbors may or may not have the same number of neighbors as you do. That is, each participant may have a different number of connections.

If you have any questions, please raise your hand and an experimenter will approach you.

Next

50.0% of your neighbors chose A in the previous round.

Press next to continue.

Next

You are currently in round 2

Choose an action from below
○ C ○ A ○ B

Hover (using mouse) and Scroll (using arrow keys) to review previous instructions

Next

Move the slider below to indicate the percentage of the participants in **the entire room** that you believe have chosen A in this round.

You will earn $0.25 if you guess within 2 percentage points (1 point in either direction) of the actual percentage.

%

0% 100%

Ending Round 2

Waiting for other participants...

Please Wait ...

0.0% of your neighbors chose A in the previous round.

Press next to continue.

Next

[…]

Demographic Survey

Please enter your age in the box below

Please select your gender from below

Continue without responding ▾

Please select your race/ethnicity from below

Continue without responding ▾

Please select your education level from below

Continue without responding ▾

Next

Thank You for Participating

You have successfully completed the experiment.

You began the experiment with a show-up pay of $10.0. Your earnings at the end of the experiment were $4.5. Your final pay amounts to $14.5.

Please wait for your number to be called by the experimenter.

References

Acemoglu, Daron, Munther A. Dahleh, Ilan Lobel and Asuman Ozdaglar. 2011. “Bayesian Learning in Social Networks” *Review of Economic Studies*, 78(4): 1201-1236.

Baillon, Aurélien, Zhenxing Huang, Asli Selim and Peter P. Wakker. 2018. “Measuring Ambiguity Attitudes for All (Natural) Events” *Econometrica*, 86(5): 1839-1858.

Benjamin, Daniel J., Matthew Rabin and Collin Raymond. 2016. “A Model of Nonbelief in the Law of Large Numbers” *Journal of the European Economic Association*, 14(2): 515-544.

Bohnet, Iris and Richard Zeckhauser. 2004. “Trust, Risk and Betrayal” *Journal of Economic Behavior & Organization*, 55(4): 467-484.

Brandenburger, Adam. 1996. “Strategic and Structural Uncertainty in Games” *Wise Choices: Games, Decisions, and Negotiations*, ed. R. Zeckhauser, R. Keeney and J. Sibenius. Harvard Business School Press.

Chandrasekhar, Arun G., Horacio Larreguy and Juan Pablo Xandri. 2020. “Testing Models of Social Learning on Networks: Evidence from Two Experiments” *Econometrica*, 88(1): 1-32.

Charness, Gary, Francesco Feri, Miguel A. Meléndez-Jiménez and Matthias Sutter. 2014. “Experimental Games on Networks: Underpinnings of Behavior and Equilibrium Selection” *Econometrica*, 82(5): 1615-1670.

Charness, Gary, Uri Gneezy and Alex Imas. 2013. “Experimental Methods: Eliciting Risk Preferences” *Journal of Economic Behavior & Organization*, 87: 43-51.

Charness, Gary and Dan Levin. 2005. “When Optimal Choices Feel Wrong: A Laboratory Study of Bayesian Updating, Complexity, and Affect” *American Economic Review*, 95(4): 1300-1309.

Charness, Gary and Dan Levin. 2009. “The Origin of the Winner’s Curse: A Laboratory Study” *American Economic Journal: Microeconomics*, 1(1): 207-236.

Charness, Gary, Michael Naef and Alessandro Sontuoso. 2019. “Opportunistic Conformism” *Journal of Economic Theory*, 180: 100-134.

DeGroot, Morris H. 1974. “Reaching a Consensus” *Journal of the American Statistical Association*, 69(345): 118-121.

DeGroot, Morris H. and Mark J. Schervish. 2012. *Probability and Statistics* (Fourth Edition). Pearson Education.

Eichberger, Jürgen and David Kelsey. 2000. “Non-Additive Beliefs and Strategic Equilibria” *Games and Economic Behavior*, 30(2): 183-215.

Eichberger, Jürgen, David Kelsey and Burkhard C. Schipper. 2008. “Granny Versus Game Theorist: Ambiguity in Experimental Games” *Theory and Decision*, 64(2-3): 333-362.

Erev, Ido and Alvin E. Roth. 1998. “Predicting How People Play Games: Reinforcement Learning in Experimental Games with Unique, Mixed Strategy Equilibria” *American Economic Review*, 88(4): 848-881.

Esponda, Ignacio and Emanuel Vespa. 2014. “Hypothetical Thinking and Information Extraction in the Laboratory” *American Economic Journal: Microeconomics*, 6(4): 180-202.

Fatas, Enrique, Shaun P. Hargreaves Heap and David Rojo Arjona. 2018. “Preference Conformism: An Experiment” *European Economic Review*, 105: 71-82.

Gale, Douglas and Shachar Kariv. 2003. “Bayesian Learning in Social Networks” *Games and Economic Behavior*, 45(2): 329-346.

Gilboa, Itzhak and David Schmeidler. 1989. “Maxmin Expected Utility with a Non-Unique Prior” *Journal of Mathematical Economics*, 18(2): 141-153.

Gilboa, Itzhak and David Schmeidler. 1993. “Updating Ambiguous Beliefs” *Journal of Economic Theory*, 59(1): 33-49.

Gioia, Francesca. 2017. "Peer Effects on Risk Behaviour: The Importance of Group Identity" *Experimental Economics*, 20(1): 100-129.

Goeree, Jacob K. and Leeat Yariv. 2015. "Conformity in the Lab" *Journal of the Economic Science Association*, 1(1): 15-28.

Golub, Benjamin and Matthew O. Jackson. 2010. "Naive Learning in Social Networks and the Wisdom of Crowds" *American Economic Journal: Microeconomics*, 2(1): 112-149.

Grimm, Veronika and Friederike Mengel. 2020. "Experiments on Belief Formation in Networks" *Journal of the European Economic Association*, 18(1): 49-82.

Heinemann, Frank, Rosemarie Nagel and Peter Ockenfels. 2009. "Measuring Strategic Uncertainty in Coordination Games" *Review of Economic Studies*, 76(1): 181-221.

Huck, Steffen and Georg Weizsäcker. 1999. "Risk, Complexity, and Deviations from Expected-Value Maximization: Results of a Lottery Choice Experiment" *Journal of Economic Psychology*, 20(6): 699-715.

Ivanov, Asen. 2011. "Attitudes to Ambiguity in One-Shot Normal-Form Games: An Experimental Study" *Games and Economic Behavior*, 71(2): 366-394.

Jackson, Matthew O. and Leeat Yariv. 2011. "Diffusion, Strategic Interaction, and Social Structure" in *Handbook of Social Economics* (Vol. 1), ed. J. Benhabib, A. Bisin, and M. O. Jackson. North-Holland.

Kahneman, Daniel and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk" *Econometrica*, 47(2): 363-391.

Köbberling, Veronika and Peter Wakker. 2005. "An Index of Loss Aversion" *Journal of Economic Theory*, 122: 119-131.

Kőszegi, Botond and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences" *Quarterly Journal of Economics*, 121(4): 1133-1165.

Levin, Dan, James Peck and Asen Ivanov. 2016. "Separating Bayesian Updating from Non-Probabilistic Reasoning: An Experimental Investigation" *American Economic Journal: Microeconomics*, 8(2): 39-60.

Machina, Mark J. and Marciano Siniscalchi. 2014. "Ambiguity and Ambiguity Aversion" *Handbook of the Economics of Risk and Uncertainty* (Vol. 1), ed. M. Machina and K. Viscusi. North-Holland.

Martínez-Marquina, Alejandro, Muriel Niederle and Emanuel Vespa. 2018. "Failures in Contingent Reasoning: The Role of Uncertainty" *American Economic Review*, 109(10): 3437-3474.

O'Donoghue, Ted and Jason Somerville. 2018. "Modeling Risk Aversion in Economics" *Journal of Economic Perspectives*, 32(2): 91-114.

Rabin, Matthew. 2013. "Incorporating Limited Rationality into Economics" *Journal of Economic Literature*, 51(2): 528-543.

Rabin, Matthew and Joel L. Schrag. 1999. "First Impressions Matter: A Model of Confirmatory Bias" *Quarterly Journal of Economics*, 114(1): 37-82.

Reiter, Andrea M. F., Shinsuke Suzuki, John P. O'Doherty, Shu-Chen Li and Ben Eppinger. 2019. "Risk Contagion by Peers Affects Learning and Decision-Making in Adolescents" *Journal of Experimental Psychology: General*, 148(9): 1494-1504.

Trautmann, Stefan T. and Gijs van de Kuilen. 2015. "Belief Elicitation: A Horse Race Among Truth Serums" *Economic Journal*, 125(589): 2116-2135.

Tversky, Amos and Daniel Kahneman. 1973. "Availability: A Heuristic for Judging Frequency and Probability" *Cognitive Psychology*, 5(2): 207-232.

Zizzo, Daniel John, Stephanie Stolarz-Fantino, Julie Wen and Edmund Fantino. 2000. "A Violation of the Monotonicity Axiom: Experimental Evidence on the Conjunction Fallacy" *Journal of Economic Behavior & Organization*, 41(3): 263-276.