Journal of Behavioral and Experimental Economics 64 (2016) 41–49

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

Logical omniscience at the laboratory☆

Ralph C. Bayer[a,*], Ludovic Renou[b]

[a] Department of Economics,University of Adelaide, 10 Pulteney Street, Adelaide 5005, Australia
[b] Department of Economics, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom

ARTICLE INFO

Article history:
Received 19 May 2014
Revised 19 May 2016
Accepted 6 June 2016
Available online 11 June 2016

JEL classification:
C70
C91

Keywords:
Iterative reasoning
Depth of reasoning
Logical omniscience
Rationality
Experiments
Other-regarding preferences

ABSTRACT

This paper investigates the ability of individuals to make complex chains of reasoning, similar to those underlying the logic of iterated deletion of dominated strategies. Controlling for other-regarding preferences and beliefs about the rationality of others, we show, in the laboratory, that the ability of individuals to perform complex chains of iterative reasoning is better than previously thought. We conclude this from comparing our results with those from studies that use the same game without controlling for confounding factors. Subjects were able to perform about two to three iterations of reasoning on average as measured by our version of the Red-Hat Puzzle.

> *The view that machines cannot give rise to surprises is due, I believe, to a fallacy to which philosophers and mathematicians are particularly subject. This is the assumption that as soon as a fact is presented to a mind all consequences of that fact spring into a mind simultaneously with it. It is a very useful assumption under many circumstances, but one too easily forgets that it is false. A natural consequence of doing so is that one then assumes that there is no virtue in the mere working out of consequences from data and general principles.* Alan Turing (Turing, 1950)

1. Introduction

Logical omniscience and rationality are two central assumptions in Game Theory. A player is *logically omniscient* if he knows *all* logical implications of his knowledge and *rational* if he chooses optimal strategies given his knowledge and beliefs. The aim of this paper is to experimentally measure the degree of logical omniscience (and rationality) of individuals, controlling for other-regarding preferences and beliefs about the rationality and omniscience of others.

All experimental attempts to measure the degree of logical omniscience (and rationality) in humans by analyzing behavior in *strategic games* necessarily conflate auxiliary hypotheses on subjects' perception of the cognitive abilities and preferences of others. Bounded rationality and other factors (strategic uncertainty, social preferences, overconfidence, etc.) cannot be cleanly separated in such experiments. This paper proposes a novel experimental design, which makes it possible to measure the degree of logical omniscience and rationality of individuals with as few confounding factors as possible.

To see that measurement without confound is difficult, consider the seminal beauty contest game (Nagel, 1995). Deducing the level of a subject's level of logical abilities from the number chosen is bound to be biased. For instance, a scholar of game theory would choose a reasonably high number if she believes that the iterative abilities of others are low, despite having the ability to iterate to the equilibrium choice of zero. After all, the optimal choice is to best-reply to one's conjecture about the choices of others, not necessarily to play equilibrium (unless one conjectures that others play according to equilibrium). Therefore, direct measures of logical abilities from observed behavior in strategic-form games are bound to be biased, as they do not take into account that play is not only a result of cognitive abilities, but also of a player's beliefs

☆ We thank the Economic Design Network Australia and the University of Adelaide for funding. We are grateful to Karl Schlag, Konstantinos Tatsiramos, and seminar and conference participants at the Australian National University, the University of Melbourne, the University of Western Australia, the California Institute of Technology, and the University of East Anglia.

* Corresponding author. Tel: +61 (0)8 83134666.
E-mail addresses: ralph.bayer@adelaide.edu.au (R.C. Bayer), lrenou@essex.ac.uk (L. Renou).

http://dx.doi.org/10.1016/j.socec.2016.06.003
2214-8043/

about the play of others. Agranov et al. (2012) show that the beliefs about the rationality of others indeed play an important role, as in their study the number of iterations performed in the guessing game varies in the expected way, when beliefs about the cognitive abilities of other players are manipulated. Disentangling own cognitive ability and beliefs is made even more difficult by the fact that not all subjects adjust their iteration depth in the same way. Gill and Prowse (2015) show that only subjects with high cognitive abilities adapt their behavior to information about the cognitive abilities of others. Alaoui and Penta (2015) develop a model of how own cognitive abilities and beliefs about the the cognitive abilities of others translate into behavior and find support for their theory using the 11–20 game (Arad and Rubinstein, 2012). Social preferences and preferences for social efficiency are additional confounding factors.

This paper acknowledges the problem of the confound and makes a methodological contribution towards solving it. We offer an experimental design that makes it possible to measure the ability of individuals to perform chains of iterative reasoning with as little confounding factors as possible, without sacrificing a game-like structure. The resulting measure of logical omniscience can be used as an explanatory variable for observed behavior in strategic-form games, which allows for an assessment of the degree to which limited cognitive abilities contribute to deviations from Nash behavior.

The experiment we designed is a variant of the Red Hat Puzzle (also known as the Dirty Faces Game), in which we control for other-regarding preferences and beliefs about the rationality of others. In the Red Hat Puzzle (RHP thereafter), a player has to determine her type (hat color) by the use of iterative reasoning. For this purpose the player can use her knowledge about the types of the other players and the other players' actions. The distribution of types determines how many iteration steps a player has to perform in order to arrive at the correct answer.[1] In its original form (as used by Weber (2001) or Bayer and Chan (2009)), the RHP suffers from the same problems as other interactive games when used to measure subjects' iteration ability. Players have to rely on the iterative abilities of other players. Therefore, not only their own iterative ability matters but also their beliefs about the ability of others, beliefs about beliefs about the ability of others, etc.[2] Social preferences might also play a role. To overcome this problem we do the following: we transform the RHP into an interactive decision problem where every "player" at each move has a unique logically correct answer. In each game, a single human player plays with computer players only.[3] The computer players are programmed to be logically omniscient, i.e. they always choose the logically correct answer. This fact is communicated to the human player. In this setup a human player, who is able to perform the necessary number of iteration steps for a particular puzzle, can fully rely on the other players' logical omniscience. Additionally, we do not have to worry about the influence of social preferences as the human player does not interact with other humans.[4] While this transformation makes the RHP a decision problem, it still remains interactive. Computer-players interact with the human-subjects in that their "actions" will depend on the action of the human subjects, and vice versa. With this procedure, we can cleanly isolate and measure the iteration ability of humans in an interactive situation by varying the type distribution within a subject.

Our experiments highlight two interesting patterns. Firstly, subjects were able to perform about two to three steps of iterative reasoning on average, more than the one to two steps typically measured in similar games without control for beliefs about the rationality of others. It is important to stress that comparisons with previous studies that do not control for social preferences or beliefs about the rationality of others are difficult. Without additional assumptions on the preferences and beliefs about the rationality of others, it is not possible to infer the ability to perform steps of iterative reasoning from observed play in strategic-form games, for example. A second result refers to learning: to our surprise, subjects did not only learn from observation (feedback). Introspection alone was sufficient for subjects to perform better when playing the same puzzles for a second time.[5] Our econometric analysis is organized around these two themes (Section 4).

This paper contributes to the large literature on iterative reasoning in games e.g., McKelvey and Palfrey (1992); Beard and Beil (1994); Nagel (1995); Ho et al. (1998); Goeree and Holt (2001); Van Huck et al. (2002); Cabrera et al. (2006), to name just a few.[6] A recurring feature of many of these studies is the use of games solvable by iterated deletion of strictly or weakly dominated strategies.[7] In these studies, the ability of individuals to perform iterative reasoning is associated with their ability to iteratively delete dominated strategies. Centipede games (e.g., McKelvey and Palfrey (1992)) and beauty contest games (introduced to the literature by Nagel (1995)) are two of the most commonly used games in that literature. However, in those games, iterating to the equilibrium might actually not be optimal for a subject. E.g, in a centipede game, a fully rational and omniscient player will pass instead of ending the game right away, as prescribed by a subgame-perfect Nash equilibrium, if she beliefs that the opponent does not understand equilibrium logic and will pass given the next move. Without controlling for beliefs about the rationality and logical omniscience of others, failure to play the equilibrium cannot be interpreted as limited ability to perform iterated reasoning.[8] Consequently, a researcher interested in the ability of humans to perform chains of iterative reasoning might underestimate the actual ability of humans when relying on choices in beauty contest or centipede games alone. The same is true, to our knowledge, for all studies of interactive games aiming to measure the iteration abilities of humans.[9]

Two closely related experimental studies are Weber (2001) and Bayer and Chan (2009). Weber implements the red hat puzzle as a dynamic game of incomplete information between two or three human players. Bayer and Chan replicate Weber's experiment and

[1] A detailed description of the puzzle will be given below.

[2] The methodology used in this paper has first been described in a conference paper (Bayer and Renou, 2007), which is based on the data from a pilot for this study. The pilot only contained a few sessions of one of the six treatments presented here. The conference paper's purpose was to describe the methodology, while this paper shows how behavior changes across the treatment dimensions.

[3] For others experimental designs with automated opponents, see Johnson et al. (2002) and Mc Kinney and Huyck (2007).

[4] Naturally, we cannot exclude the possibility that the subjects had concerns for the well-being of other persons affected by their decisions, e.g., the experimenters, other students (perhaps because the funding used in the experiment could have helped these students), etc. This is unlikely to play a major role, though. For instance, Frank (1998) and Fleming and Zizzo (2015) have not found evidence of altruism towards experimenters.

[5] A similar observation is made in Weber (2003).

[6] We refer the reader to chapter 5 of Camerer (2003) for a survey of this literature.

[7] Note that the solution concept of iterated deletion of weakly dominated strategies requires more stringent conditions than common knowledge of rationality (see Brandenburger et al. (2008)).

[8] The same is true for beauty contests where a logically omniscient player chooses the number corresponding to one more iteration step than he believes the others are able to perform. Failing to choose the equilibrium number is not necessarily a sign of limited iterative ability.

[9] Gneezy et al. ([2007,2010]) and Dufwenberg et al. (2008) use a version of the game "Nim" to study if and how humans learn backward induction. Since there players have (weakly) dominant strategies, this zero-sum game can be used to infer the depth of counterfactual reasoning from the steps of backward induction performed, if one accepts the auxiliary hypothesis that it is common knowledge that nobody deliberately plays weakly dominated strategies.