Universitat Pompeu Fabra Barcelona
Department of Economics and Business

Economics Working Paper Series

Working Paper No. 1539

Inspired and inspiring: Hervé Moulin and the discovery of the beauty contest game

Rosemarie Nagel
Christoph Bühren
Björn Frank

Updated version: November 2016
(October 2016)

Inspired and Inspiring:

Hervé Moulin and the Discovery of the Beauty Contest Game

by Rosemarie Nagel* (Universitat Pompeu Fabra, ICREA, BGSE),

Christoph Bühren and Björn Frank♦ (University of Kassel)

Abstract

We draw an unusually detailed picture of a discovery, the beauty contest game – with Hervé Moulin as the center of the initial inspiration. Since its inception, the beauty contest game and the descriptive level k model has widely contributed to the growth of experimental and behavioral economics and expanded also to other areas within and outside of economics. We illustrate, in particular, the recent interaction between macroeconomic theorists and experimenters, who independently had worked on the puzzles and consequences due to beauty contest features. Furthermore, we introduce a new variety of the two-person beauty contest game with two different payoff structures that leads to different game-theoretic properties unperceived by naïve subjects and game theory experts alike.

Keywords: Keynes, Beauty Contest Games, Guessing Games, History, Level k, Micro-, Macro-, Neuro-Economic Experiments

JEL codes:C9, D84, D87, E12, N1, N80

---

* rosemarie.nagel@upf.edu, ICREA, Universitat Pompeu Fabra, Department of Economics, BGSE, Barcelona, Spain. Nagel thanks the grant ECO2014-56154-P of the Spanish Ministry of Education; Generalitat de Catalunya, and Barcelona Graduate School of Economics.

We thank Carlos Alos-Ferrer, Larbi Aloui, Ayala Arad, Anna Bayona, Jess Benhabib, Antoni Bosch, Andrew Colman, Eddie Dekel, John Duffy, Ernan Haruvy, Nagore Iriberri, David Freeman, Williamine Kets, Felix Mauersberger, Shabnam Mousavi, Kristoffer Nimark, Alvin Roth, Ivan Werning, Michael Woodford, and one anonymous referee for helpful comments. Special thanks are due to Alain Ledoux, the editor of this special issue Jean-François Laslier, Martin Hellwig, Tore Ellingson, and Eric Van Damme for their personal accounts of their "discovery" of the beauty contest game.

buehren@uni-kassel.de, University of Kassel, Institute of Economics, Nora-Platiel-Straße 4, 34127 Kassel, Germany, Tel.: +49-561-804-7267, Fax: +49-561-804-3083

♦frank@uni-kassel.de

1. Introduction

Keynes' (1936, p. 156) stated with his famous metaphor of stock market investment as a newspaper beauty contest:

> *"professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; […] It is not a case of choosing those [faces] which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practise the fourth, fifth and higher degrees.*" (Keynes, 1936, ch.12.V).

This became known in the experimental economics literature as the guessing game or the (p-)beauty contest game. While Keynes' beauty contest seems to capture a reasoning process of higher order beliefs, it could as well be misleading: in the mind of the observer the face chosen based on the average opinion is the same as by "third degree". Of course Keynes was aware of this outcome equivalence and left it to future generations to find the right game to visualize the separation of different beliefs in actual choices.

Guessing or p-beauty contest games, where players have to guess 2/3*average of all chosen numbers, indeed generate different outcomes, clearly distinguishing higher from lower order beliefs. They constitute one of the most important tools for investigating human bounded rationality and higher-order beliefs in relation to game-theoretic concepts such as common knowledge of rationality, rationalizability, and Nash equilibrium reached through iterated elimination of (weakly dominated) strategies or fixed point arguments. Experiments on the games brought to light the lack of accordance between human behavior and these theoretical concepts, suggesting alternative cognitive procedures, level k and cognitive hierarchy models (Nagel, 1995, Camerer et al., 2004; for surveys see Camerer, 2003, Crawford et al., 2013).

We maintain that Hervé Moulin remains in the center of the origin of the game in economics for the ongoing chain of evolution in experimental and behavioral economics and other areas in economics and other disciplines:

> *"How realistic is the strategic behavior implied by the concept of sophisticated equilibrium? It presupposes that each and every Player either computes all dominated strategies …[This] does not cause much trouble. […] [T]he successive elimination can be arbitrarily long. In practice, most players do not perform the elimination forever out of fear that other players are not rational enough to do so. This is what experimental evidence of the "guess the average" game suggests. [...] [Y]ou do not expect that all of them will perceive the geometrical shrinking of strategy sets."* (Moulin, 1986; see also appendix I).

But who invented the game? Was it inspired through Keynes's passage? Let us start from the end of the story and trace back the path of the origin and evolution. In a recent online beauty contest experiment with over 6,000 chess players, subjects were asked to pick a number between 0 and 100, where the winning number is two-third of the average of all picks. Bühren and Frank (2012) found that chess players, and even chess grandmasters, act similar to regular student participants in experimental beauty contests in the first round (with spikes at $50*2/3^k$, for k=1,2,3, and infinite, called choices according to the level k model, with an average 32.15) and over time. Guesses of better chess players were a bit closer to 0 and a bit closer to the winning number, but the associated effect sizes were very small.

While this finding is in line with most experimental research on chess players (who turn out to differ from other humans in exactly one aspect, namely their chess playing strength), a major surprise outcome was one participant, Alain Ledoux. He claimed to be the inventor of the beauty contest game and as evidence sent us two articles (Ledoux, 1981, and Paclet, 1983, see appendices II. and III.). Thus, we (Bühren and Frank) realized that the roots of the beauty contest experiment were, indeed, partly lying in the dark. This set us on a quest to trace back to the beginning of the beauty contest game. We discovered that the origins of the two articles send by Ledoux did not link to Keynes (1936). Instead it contained many relevant experimental methods such as games appearing repeatedly in the same newspaper, as well as behavioral concepts such as level k, related to the reasoning mentioned in the quotes above, which were later independently (re)invented by experimental and behavioral economists.

In this paper, after reporting the roots of the origin of the game we discuss experimental and behavioral literature related to the beauty contest game. We accentuate the emerging impact of level k related models on micro theory, especially on the epistemic game theory. We bring forward a recent fruitful interaction between microeconomic experimenters and macroeconomic theorists through the channel of the beauty contest game and limited reasoning models. Macro theorists have, long before experimenters did, discussed independently the effects of the beauty contest game originating from Keynes (1936).

We then introduce a new two-person game, which can be traced back to the original articles of Ledoux (1981) and Moulin (1986). We implement two different payoff structures. One is a tournament payment and the other one a continuous payment with respect to the distance of a choice from the winning number; both with game theoretic properties different from the original games. Interestingly, neither our naïve subjects nor (to a lesser degree) our expert subjects such as game theorists, indicate awareness of these fine theoretical differences when playing against boundedly rational humans, leading to costly welfare implications.

The rest of this paper is structured as follows. Section 2 unearths the inspiration of Hervé Moulin, reveals the creative process of Moulin's discovery of the game's potential, and outlines the findings of the seminal experiments (re)-played around ten years later. Section 3 shows how the game transformed into a lab experiment which sets the foundation of a specific non-equilibrium model, the level k model, and section 4 presents its multiple impacts within the economics. Section 5 introduces a new two-person game; section 6 concludes and acknowledges Moulin for his creativity and inspirational role.

2. Inspiration and creativity

What can economists do to retain their creativity, i.e., to make sure they do not always trot along the old sterilized paths? While interdisciplinarity is a popular answer, Hervé Moulin has taken a different route beautifully described and praised by Falkinger: Instead of interdisciplinarity, Moulin practiced *indiscipline*, a critically controlled lack of discipline, allowing the researcher to "leave the four walls of science" (Falkinger 1988, p. 10; translation is ours), but then, after frontier crossing and letting himself drift away, returning to the desk and reflecting on whatever he has encountered [while roaming outside the walls].

It is typically difficult to deconstruct the creative process that has triggered a new idea. What we do in this paper is to tell a story of a scientific discovery in more than usual detail, where Moulin is both the one being initially inspired as described by Falkinger (1988), and then, crucially, inspiring others.

"Jeux et Stratégie" (1980-1990) was a bi-monthly popular magazine devoted mainly to strategic board games such as chess that also covered card games, mathematical games, and comic strips. In 1981, it arranged a sizable readers' competition consisting of mathematical puzzles as well as problems from games such as chess, bridge and go. Ledoux (1981, see also appendix II) reports on almost 15,000 participants, of which 4,078 *ex aequo*, leading to a playoff for choosing ultimate winners. All first round winners received a letter with new puzzles. To avoid another round with thousands of winners, chief editor Alain Ledoux invented in the last question of this letter what is today known as the beauty contest. Readers were asked to state an integer between 1 and 1,000,000,000, where the winners will be the ones closest to two third of the average, in decreasing order. The average turned out to be 134,822,738.26, which is 13.48% of the maximum number and unusually small compared to beauty contest lab experiment modal results. 350 winners "with reasonable numbers" were announced in the resolution article of the puzzles with prizes for the best 250 winners ranging from a computer Victor Lambda as the first prize, Skirrid, Zahkia, Quad-omnos games and so on. The low resulting average may be explained by a double pre-selection: Participants of the tie break were those readers of a strategic board game magazine who won the readers´ competition[1]. Indeed, the closest result can be found in Bosch-Domènech et al. (2002) from experiments with economists in conferences and seminars and advanced economic students (average 18.98).

Ledoux (1981) only briefly reports on the results without any frequency figure by pointing out that by "good sense" nobody should report an "important" number but instead the lowest number 1. However, he continued, this is not a very reasonable choice, except when one expects only one other player as an opponent - *et voila*, our two-person game. In contrast, he calls the non-zero average a "logical result".

---

[1] The effect of the double pre-selection is apparently not trivial, as Alain Ledoux submitted an even lower number than (2/3)*13.48 in the first round of Bühren and Frank’s (2012) experiment.

In 1983, in the same French magazine the same game was used as a supplement question with a similar motive to extract hundreds instead of thousands of winners who correctly answer the puzzles in the main phase of the competition. The (new) readers were informed that the same game already appeared in 1981. This is an early instance of an experimental replication, and likely the first repeated game of this kind. In this second round of the game, the best winner chose 33,289,564, thus about two steps below the winning number of the "first round" and won a Console Colcovison. Many prizes were also given to subsequent winners. Another participant from a different subset of all participants had the best winning number about 1 step from the 1981 results (67,373,773) and won a trip to Mexico (see appendix III).

Unlike Alain Ledoux, who merely considered the outcomes of the 1981 game as a logical result, the mathematician Phillipe Paclet (1983) spelled out several possible approaches to the game by analyzing the responses from the 1983 contestants (see his article in appendix III. and his recent personal comments in appendix V.). Paclet started with the analogy that a computer might not be superior to humans when reasoning about the average. He continued with outlining what is now called the level k behavioral model by assuming 2999 monkeys playing the game with one human to justify the answer 33,333,333, where the monkeys would play randomly with half a billion as the average. However, he immediately reasoned that one can doubt to be able to find just one answer with this procedure but instead a whole sensible range of answers can arise.[2] Clearly, numbers above 666 million (2/3 of the maximum number) should not be chosen, and thus one "mechanically" arrives at 1 through repeating multiplications by 2/3. But this cannot be reasonably expected because of the existence of particular players who chose based on different reasoning. For instance, there will be players who simply choose an arbitrary telephone number, or those who “artificially” increase the average by playing the highest number, which will be advantageous for a group of subjects who coordinate their entries, and indeed such groups existed among the participants.

The main part of Paclet’s argument is a discussion of Bayesian reasoning by specifying different distributions of answers and best responses, using the actual numbers. He also mentions that some players ignored or might not have known at all the first round results, which simulates new entrants. He jokes about the winner with the number 67,373,773, who apparently likes the numbers 3 and 7. He ends the two-page article with a wish to have received the "secret reflections" of the participants of the contest and that they should submit them next time. But he immediately reverts that there cannot be another round of this game. Instead the readers should propose other games for the subsidiary question. Replete with game theoretic insights and methods, the only technical term from game theory in the entire article is "equilibrium".

Interestingly, every little detail mentioned in Paclet’s article, despite the high heterogeneity of behavior and arguments, was (re)discovered independently in new experiments in the lab and field, including

[2] Paclet did not, however, mention or realize that the winning numbers of his experiment where one or two levels below the winning number of the previous winning number of Ledoux (1981).

Paclet's wish for knowing what went on in the heads of the participants during the game coming true with fMRI images 25 years later (Coricelli and Nagel, 2009) and recording of comments by contestants in the lab and new newspaper experiments (Bosch-Domènech et al., 2002).

One of the readers of "Jeux et Stratégie" was Hervé Moulin.[3] He recognized that Ledoux´s game is an example of successive elimination of dominated strategies. In chapter 4 of the second edition of his game theory book, entitled "sophisticated and perfect equilibria", Moulin included an adaptation of Ledoux's game: Choose a number between 1 and 999 where the winner is the closest to 2/3 of the average[4] (Moulin, 1986, p. 72, see appendix I). As we will discuss in sections 3 and 4, this game inspired an extensive line of research. Note that Moulin allowed in his definition of the game only one winner or those who tie. He could have gone into the other extreme suggesting payments according to the distance of the target number from the same instructions to pay all "reasonable numbers". Moulin (1986) also realized that this equilibrium is hardly achieved by human subjects (see quote in the introduction), "[…] the winning guess usually lies between 100 and 200 [when the maximum is 999]." This empirical result indicates that he played the game with advanced undergraduate or doctoral students (see figure 1 below and Bosch-Domènech et al., 2002) where level k choices are less prominent (see section 5). It also adds to the evidence from the field produced by two newspaper articles mentioned above, and to the independent replications that started to be generated about 10 years later in labs, classrooms, and with newspapers.

Moulin did not mention explicitly the two-person case, which might not have belonged to his chapter of "sophisticated and perfect equilibria", as it seems to be a quite unsophisticated equilibrium, in (weakly) dominant strategies, similar to the equilibrium in the prisoner's dilemma game. We will expand further on this line of thought in section 5.

We observe that Moulin's endeavor manifests a complete exercise of Falkinger's concept of indiscipline. That one must, through terminology and outlet, be able to reach the fellow scientists from one's discipline. Our readers might wonder why it took 25 years until the original source was re-discovered. The next section provides a simple answer.

3. From the field and theory to the lab

Since its discovery by Hervé Moulin in the early 1980s, the beauty contest game was used in a few advanced undergraduate and PhD economics classes as a demonstration experiment (for failures) of common knowledge of rationality, rationalizability, dominance solvability, etc. In November 1990, at the London School of Economics, Tore Ellingsen presented a one-shot guessing game in his master Industrial Organization class which I (Rosemaire Nagel) attended as a visiting PhD student. I chose 22,

---

[3] We do not know if he also took part in the competition.

[4] In a personal email Ledoux (see appendix V) expressed his surprise that even the parameter 2/3 has survived until today which he had chosen more or less arbitrary.

a rounded number according to 50*2/3*2/3 without even contemplating further steps. The winning number turned out near 20 and I did not further think about this game.

Shortly thereafter, I, together with a few other students from Tore's class encountered the game again in January 1991, in Roger Guesnerie's Phd game theory class. This time around, I chose a number a bit lower than the previous winning number (near 20) and won; the chosen numbers were written on the blackboard with some choices being indeed 22, 33, or close to these. When Guesnerie repeated the game for the second time, I won again with a number near 9, which is found approximately by taking 2 steps from the winning number of the previous period. Thus, I had gained experience from my first exposure, then by playing against a mixture of experienced and inexperienced subjects, and finally repeating the game against experienced players.

There I realized that a simple procedure can classify the behavior in this game. Would I have ever started to work on this topic have I never personally played the game or been triggered by winning? Doubtful. I asked Guesnerie to give me the data, analyzed them, and also ran a pilot with undergraduate students at the London School of Economics. To be a subject in one's own study (as many medical researchers do) together with being repeatedly exposed to an idea offers one a deep insight into the beauty and potentials of the subject matter in a way that detached study might not offer[5]. Not to be overlooked or forgotten are the many random draws, pain, and luck that accompany a researcher along the path of discovery, especially when one sees something that others do not see.

Remember that the beauty contest game had been around for some years, but nobody viewed it useful as a proper lab experiment. First of all, graduate students were not acquainted with the idea of lab experiments[6] or considering themselves as experimenters. Furthermore, theorists only noticed the out-of-equilibrium behavior and an obvious convergence over time, without any clear structure behind that. Thus, for taking interest in the beauty contest game most comments I received were along the lines of, "There are other more interesting games, like the centipede game" (Ken Binmore), "Don't waste your time", or "How can you do a thesis with such a game?"[7]

These reactions did not give me cold feet. Intrigued by playing the game at Tore's and Guesnerie's classes, who used the beauty contest game as a demonstration experiment to show the failure of rationalizability, and the convergence over time, I started with lab experiments strongly supported by

[5] As pointed out by Bergstrom (2003), Vernon Smith is a prime example. Experimentalists are usually well equipped to deal with counterfactuals, however, we will never know which turn Vernon Smith's career would have taken had he not been a student subject in one of Edward Chamberlin's classroom market experiments.

[6] This is still true today for many economic (graduate) students. I had participated in many lab experiments during my undergraduate study in Bonn, and during lectures, especially by Eric Van Damme.

[7] E. Van Damme personal accounts: "This [knowing about the game] must have been around 1980….it did not cross my mind to do experiments with this game at the time. As we now know, nobody thought about it at the time…Later the rationalizability concepts drew much attention, but even then at first nobody did serious experiments. When your work appeared, I guess that many of us might have regretted not having seen the goldmine in front of us." (see appendix V). When Dale Stahl came to Bonn in 1992 to present his work, published later in Stahl (1993), it was due to Eric's suggestion that he approached me for an exchange about my experiment.

my advisor Reinhard Selten (see also Coricelli and Nagel, 2010) to find evidence of the $50*p^n$ model in the reasoning of human subjects and convergence patterns over time.

Previously, Mitzkewitz and Nagel (1993) study ultimatum games with incomplete information and classify players into anticipation types (e.g. those who anticipate the acceptance levels of responders under expected fairness considerations). However, higher order beliefs could not be investigated. The beauty contest game by contrast provided an ideal structure for studying the limits of common knowledge of rationality, with a new kind of anticipation, later known as level k.

In addition to simply asking for choices in my pilot study, I invited my subjects to comment on their choices. To learn how to classify the behavior based on these comments, I contacted a group of psychologists at the London School of Economics[8]. Thus, the "secret reflection" Paclet (1983) had wished for came to life. Unlike most of the follow-up models by other researchers, I abstained from non-degenerate types, i.e., a probability distribution over different types, instead I just assumed that all others choose one level lower than oneself. I was guided, maybe too strongly, by the fact that only one subject in my pilots mentioned a probability distribution over those who choose 33, 22, 14 and alike. However, also in later lab and newspaper experiments less than 5% submitted type distributions, when possible, in xls spreadsheets with best replies to distributional types.

To check the statistical validity of the model, I tested whether the choices were significantly close to the theoretical values $50*p^k$, for k=0,1,2,3, or fell in the interim intervals, and compared them also with the dominance solvable equilibration structure. Reinhard Selten supported that simple idea, however, he suggested taking the geometric instead of the arithmetic mean to construct interim intervals to capture the geometric decrease of the level of reasoning. This "casual observation" (Camerer, 2003) was later confirmed by more sophisticated econometric models using mixture models (e.g. Stahl 1996, Camerer and Ho 2004, and Bosch-Domènech et al. 2010)

A fellow PhD student Klaus Kultti concurrently at the London School of Economics in 1991 pointed at a similarity of the game (maybe also of my reasoning process) with Keynes' (1936) beauty contest mentioned in the introduction.

However, Selten was against the name "beauty contest" as Keynes' contest has multiple equilibria (as the famous Van Huyck et al., 1991, coordination games) and the types are reduced to level 0 (random types, focal choice) and level 1. Thus, we named it "guessing game" while referring to Keynes (1936) in Nagel (1995). Keynes's quote, above, explains best the analogy of finding regularities, and at the same time points at the difficulty of making predictions in real forecasting situations, intertwining multiplicity of equilibria and multiplicity of bounded rational reasoning.

---

[8] The editor of Nagel (1995) requested to delete the classification system. Only 7 years later, the referees of Bosch-Domènech et al. (2002) suggested to classify all comments submitted for the newspaper experiments.

In 1994-1995, when I was in Pittsburgh as Al Roth's post doc, John Duffy[9] insisted calling the game "Keynes' beauty contest" in Duffy and Nagel (1997), as did independently Ho et al. (1998), with "p-beauty contest game", where p=1 mirrors Keynes' original game. Al Roth suggested including "unraveling" in the title of Nagel (1995) in analogy to his matching market observations.

In the same year, two more important incidents led me closer to the origins of the game. My friend Oliver Schulte, a graduate student in computer science and philosophy in Carnegie Mellon (now computer science professor in Simon Fraser University) pointed out a then still unknown source to me, the "guess the average" game in Moulin's book (1986), which Oliver used for teaching to computer science students. Later, in the 1995 public choice conference, I met Hervé Moulin and told him about my research. He then referred me to *Pour la Science* (French version of the *Scientific American*) as his original source. Neither Moulin's nor my search in the logical puzzles sections of all volumes of *Pour la Science* from 1979 to 1981 led to discovery of relevant material.

As it turned out, we had to wait another 14 years, until 2009, for another strike of chance to find the origin: Fishing in the international subject pool of online chess players, eventually revealed the original inventor of the game, Alain Ledoux, who gave us the initial source. It was *Jeux & Stratégie* with Ledoux as one of the founders, edited by *Science & Vie,* a journal with the same content coverage as *Pour La Science*, "which might have explained Moulin's confusion" as Ledoux noted in an email (see appendix V). Thus, we both had looked for the needle in the wrong haystack.

On top of all these unlikely chain of events which led to unearthing the original inventor of the game, there is yet another fortunate "after play" that materialized through the editor of this special issue, Jean-François Laslier. After reading the initial submitted version of this paper, he wrote to us about how vividly he remembers playing the game and the original source "Jeux & Estratégie" from Moulin's class in ENSAE (Ecole Nationale de la Statistique et de l'Administration Economique), a French "Grande Ecole" in 1981-1984. He also added a comment about how Moulin's teaching had influenced his later research guided by Gilbert Laffond, who was Hervé's first PhD student (see also appendix V).

4. Related experimental, behavioral, and theoretical literature

Since the 1980s there have been many new attempts to model bounded rational behavior in two streams of the theoretical literature, reasoning based on rationalizability and Bayesian game modeling, and (evolutionary) learning. In the next two subsections we discuss the experimental and theoretical lines of research, respectively, related to beauty contest games and the level k model.

[9] Duffy referred Nagel (1995) to Simonson (1988) and Frydman (1982), who discuss Keynes' (1936) beauty contest game in macro-labor contexts (see also discussion in section 4.2).

4.1. Experimental and behavioral literature

4.1.1 Level k related models and experiments

In this subsection we refer to the beauty contest experiments and the level k models arising directly from experimental data. As stated in section 3, the basic level of reasoning model (later called level k), an iterated best reply model with degenerate prior beliefs and choices, was inspired by Nagel's participation in classroom experiments. This procedure specifies very simple ideas which might have disciplined subsequent similar modeling choices (see especially the contrasting suggestions by Paclet, 1983, and Stahl, 1993, see also section 4.2.1.) using (also) probability distributions over different (level k) types:

1. A level 0 player plays in beauty contest games either the focal point 50 or a random number with an average of 50 without game form recognition, that is without (any) understanding of the rules of the game (Chou et al., 2009).
2. Level 1 player gives the best response to level 0, as if playing against nature.
3. Level k (k>1) best responds, assuming a degenerate belief and choice that all others are level k-1 players. These players use theory of mind unlike level 0 or level 1 types (Coricell and Nagel, 2009, see also section 4.1.5.). And there are those players who play the equilibrium strategy, i.e., the limit of the iterated best reply process (assuming common knowledge of rationality).
4. Errors are introduced through some randomization around the theoretical level k choices.
5. Steps 1-4 can also be applied over time, with the average or winning number of the previous period being level 0 (see e.g., Nagel, 1995, Stahl 1996, and Camerer and Ho, 1998).

This model provides a bridge between irrational or random behavior and equilibrium outcomes. It therefore departs in several aspects from the theoretical literature of (bounded) rational behavior. The starting point level 0 is guided by heuristic principles (e.g. random, focal, reference points as signals, etc.) and not (necessarily) by dominated strategies; individuals have a simple model of others with a degenerate prior and choice, believing that all others are just one type below them, which may be inconsistent with actual distributions. A level k player ignores types that use higher levels than k-1. No equilibrium is calculated, but it might be reached in the limit by experience or in the mind of the player.

The experimental data in many different games shows that typical responses belong to reasoning levels 0 to 3 with some noise. These kinds of stopping rules have provided the greatest challenge and indeterminacy for theoretic modeling, while level 0 has created an a priori indeterminacy or freedom for experimenters which can only be resolved by experimental evidence. Remember, also in Keynes's metaphor, any woman can potentially be selected by the average, or Schelling (1960) discussing focal points with a need of empirical foundations. A careful discussion on this most problematic modeling technique can be found in Heap et al. (2014) and Crawford's comment on Heap et al., studying various forms of coordination games. The basic conclusion about level 0 reasoning we draw from all these works is that it remains a challenge to provide a general theory for level 0 modeling, which cannot be done without empirical methods.[10]

The level k model is mainly a descriptive model. However, it can also be useful as a predictive model: if one can predict or conclude that level 1 choices are close to equilibrium then one can more likely expect (near) equilibrium choices, while if level 1 choices are far away, then (first period) choices will more likely be far away from equilibrium; learning depends on payoff realization for different levels and whether low level players die out (see e.g. Nagel, 1995). Bayona et al. (2016) show in supply schedule experiments that equilibrium behavior is attained when subjects have non-correlated signals about costs between different players, since level 1 choices are near equilibrium. In the treatment with non-correlated signals behavior remains far from the equilibrium, which requires higher order beliefs, and naive behavior, with high enough payoffs, does not die out. Most subjects simply ignore the correlation because of cognitive difficulties and play as in the uncorrelated case.

Most subsequent level k models differ in small details from the original one mentioned above. Stahl and Wilson (1994), based on Stahl (1993, see also section 4.2.1.) and Nagel (1993), restrict types to level k=0, 1, 2, with k=2 players best responding to a distribution of level 0 to level k-1 players, and additional Nash equilibrium types. They provided the first mixture model technology in experimental economics by classifying players into distinct types given an error structure. Subjects participate in a sequence of different 3x3 normal form games without feedback, as in Rapoport et al. (1976). Stahl and Wilson (1995) contain level k players as in Nagel (1995) and additionally "worldly players" who are Bayesian players best responding to a distribution of level 0 to level k distributions. Stahl (1996), however, analyzed Nagel's data without such Bayesian players, and for the analysis of behavior over time also introduced a horse race between different learning models, including level k learning. Indeed, Stahl (1998) convincingly argues that integer steps of level k thinking seem a more plausible format than continuous steps for Nagel's data.

Haruvy et al. (1999) and Costa-Gomes et al. (2001) departing from Stahl and Wilson (1994, 1995) add several other types of players based on payoff outcomes analyzing normal form experiments. Colman et al. (2013) discuss a large set of bounded rational strategies for stag hunt games such as team reasoning, social projection theory, Stackelberg reasoning and cognitive hierarchy theory come to the conclusion that players use several of those strategies before making their choice.

Camerer et al. (2004) propose an important alternative to level k reasoning, the cognitive hierarchy model (note that level k is not a special case of this model). Cognitive hierarchy is a one parameter

[10] This discussion is also reminiscent of the problem of the formulation of reference points in Kahneman and Tversky (1979) and related literature.

model with a Poisson distribution, specifying the proportion of players stopping after each step. Each type k best replies against a perceived distribution of all lower level reasoners. They find that the parameter 1.5 is a reasonable parameter for a wide class of experiments, recovering in particular the spikes in beauty contest games.

Bosch-Domènech et al. (2010) develop a mixture model in which the number of level k types, the means and variances of each type are estimated from the beauty contest data, instead of predetermining them as in the previous models. In fact, the actual frequencies for each type varies most across the different subject pools and less with respect to type specifications. Camerer (2003a, p. 255) argues that the responsive sensitivity parameter of the quantal response equilibrium (McKelvey and Palfrey, 1995) can be translated into the idea of iterated elimination of dominated strategies. Indeed, Breitmoser (2012) finds evidence for quantal response equilibria as well as noisy introspection models (Goeree and Holt, 2004) in the data of Bosch-Domènech et al. (2002) estimating noise parameters for each level k, as Camerer (2003a) suggests. In none of these models it is specified that level k players reason about higher level players as proposed by Stahl (1993).

Since the late 1990s experiments on beauty contest games and normal form games were inspired by testing the level k and related models, (see surveys by Camerer, 2003, Crawford et al., 2013, and Crawford, 2013). These models were validated as descriptive models in many other experiments that do not contain the dominance solvable structure, e.g. matching pennies[11] and entry games (Camerer et al., 2004). Moreover, the behavior in incomplete-information games as in private or common value auctions (Crawford and Ireberri, 2007) and asymmetric toehold common value auctions (Georganas and Nagel, 2011) are explained by level k. Naive players (level 0) choose according to simple heuristics of truth telling, the value as the bid, instead of random behavior, which Crawford and Ireberri (2007) also discuss. These players do not even need to exist in the data except in the mind of the decision maker.

Camerer (2003a) dedicates an entire chapter to dominance solvable games. Most of the games from the early 90ties are either mixed motive games which might induce fairness considerations or have inefficient equilibria as the centipede game. The behavior is classified by the procedure of iterated elimination of dominated strategies, typically with few possible steps reaching equilibrium. Indeed, also Camerer and Ho (1998) apply this kind of reasoning in the p-beauty contest games for their first period data instead of starting at the focal point 50, maybe to avoid an additional parameter for level 0 choices. For further discussions on many different micro experiments related to level k reasoning see Crawford et al, (2013) and Crawford (2013)

[11] Around 1995, Martin Hellwig emailed to Nagel his submitted back cover in the Journal of Political Economy (JPE) 1992, which JPE called "Games with Asymmetric Intelligence". A clever school boy outsmarts his classmates with level k reasoning in repeated matching penny games (“the purloined letter" by E.A. Poe, 1845, see appendix IV). Nagel (1998) comments on this quote. See also Hellwig’s personal comment in appendix V.

A great breakthrough in extending level k reasoning to complex games is introduced by Arad and Rubinstein (2012). They experiment on the Blotto game (Borel, 1921), in which two players simultaneously distribute a fixed number of “troops“ across 6 “battlefields“. The winner in a field is the person with the higher number of troops in that field; in case of a tie no one wins; each player wishes to maximize the number of wins. The authors extract decision procedures involving multi-dimensional iterative reasoning from the decisions of the subjects. They consider three dimensions (features) of strategies: 1. In how many fields to concentrate, i.e. to invest most of the resources, 2. Whether to fill “disregarded“ fields with 1 or 2 troops. 3. Into which fields to assign a large number troops (first, middle or last location). In each dimension, L0 chooses the intuitive category, L1 chooses a category that properly responds to L0 within the dimension, etc.

4.1.2 Elicitation methods to uncover different level k types

A large literature uncovers the reasoning processes mentioned above through various (new) choice elicitation methods. Costa-Gomes et al. (2001) use mouse tracking technique by hiding the information of payoffs in normal form games behind boxes which subjects need to click to receive the information. The authors specify possible look-ups-ex-ante rules. Brocas et al. (2014) find evidence for level k thinking in betting games when using this technique. Haruvy (2002) proposes direct elicitation of beliefs to support level k behavior, additional to actions. He discovered that explicit beliefs exhibit level k patterns. The eye tracking[12] results of Müller and Schwieren (2011) suggest that subjects of beauty contest experiments actually think (or look) more steps ahead than their chosen numbers reveal.

Kneeland (2015) implements a new experiment based on a ring structure which does not require the experimenter to specify level 0 players. Friedenberg et al. (2015) distinguish "bounded reasoning about rationality" from "bounded reasoning about cognition," where a subject is cognitive if he has some "method" or "theory" as to how the play the game. A subject who is rational is obviously cognitive (rationality is one possible "method"), but he may be cognitive and irrational (e.g., follow a rule of thumb). Friedenberg et al. (2015) develop a novel identification strategy to disentangle these two approaches on Kneeland's (2015) ring games. A significant percentage of the subjects have a gap between reasoning about cognition and reasoning about rationality. In particular, about half of the subjects that reason only one or two rounds about rationality are identified as having higher levels of cognition.

Agranov et al. (2015) ask subjects to continuously provide their choices in beauty contest games for a given time interval, with the payoff being determined by one randomly selected second and the choice at that time. They receive clear-cut differences between random choice patterns and iterated best reply patterns. Arjona et al. (2016) ask one set of subjects for their choices and another set about their guesses of the distribution of choices in an Arad and Rubinstein (2012) modified two-person 11 to 21 games, in

[12] An eye tracker measures where people look.

which a bonus payment is given to the player whose choice is by R=3 lower than his opponent choice, and thus they distinguish more clearly between level k choices and noisy choices than when R=1. About 74% of the guesses are level 0 to 3 choices.

Burchardi and Penczynski (2014) use two-person team beauty contest games in order to evoke written communication between team members. Sbriglia (2008) asks only the winner of a round to explain his strategy to other players which expedites convergence since players best respond to this message. Fragiadakis et al. (2013) play two-person beauty contest games as in Costa-Gomes and Crawford (2006) and ask subjects to recall their choices or play against their own previous choices from memory. They find that the payoff maximizing actions are more likely chosen by those who use level k driven strategies, compared to those who behave according to other thinking patterns. Niederle (2015) uses the beauty contest - level k combination to critically discuss good experimental design techniques.

4.1.3. Convergence over time

Another important question in beauty contest experiments is related to convergence over time, what speeds up learning (e.g. team behavior vs. individual behavior, Kocher and Sutter, 2005), or which learning models best explain the behavior (e.g. Nagel, 1995, Stahl, 1996, Camerer and Ho, 1998, Nagel, 1998, Camerer, 2003). Hindrance of learning appears when new players are inserted every 4 periods (Slonim, 2005) as the average returns back to first period averages every 4 periods.

Many experiments show that if naïve players do not die out, level k reasoning is not increasing over time, using the average of the previous period as level 0 (e.g. Nagel, 1995, Camerer et al., 2004). When naïve players have no influence as in Duffy and Nagel (1997) in 2/3 median games k increases over time. Weber (2003) finds that even if players are given no feedback about behavior of others, choices decrease over time within the same game. When the winner is paid according to his choice instead of a fixed price as in Nagel (1998), the average decreases much slower.

4.1.4. Related macroeconomic experiments

This interest on convergence over time finally brought together macro and micro experimenters, to whom the elements of beauty contest games are of essential importance (see e.g. Hommes et al., 2005, on forecasting games). Duffy (2016) provides a survey including some of these kinds of experiments among a wide range of macro experiments.

Over several years, the many studies on learning in beauty contest games failed to obtain instantaneous equilibrium play. Benhabib et al. (2016) propose a new target, with the choice of player i being closest to "2/3*average+ε^i," where ε^i is an idiosyncratic private signal, independently drawn from a normal distribution $\mathcal{N}(0, \sigma^2)$, which is common knowledge. This game stems from a simplified model of a monopolistic competition general equilibrium model in which firms receive exogenous signals from consumers, also called sentiments (see Benhabib et al., 2015). Choices can be selected from all real numbers (thus also negative numbers are allowed). The game theoretic solution is to play the private signal ε^i, as the mean choice will be zero and thus the choice i equal to the signal. The signal becomes an anchor which indeed is the modal choice, and the average signal is close to zero. Thus, the level k reasoning has been finally reduced to one level: all have to choose their signal, a clever reference point system, and thus experts and naive subjects choose alike and obtain highest payoffs, given that payoffs are paid according to the distance between own choice and target (see also the experiment in section 5).

Even when subjects get correlated noisy signals, behavior is closer to equilibrium than in the original beauty contest game, despite the theoretically difficult signal extraction problem. The reason is that those signals are better anchors than signals arising endogenously in the head of the subjects in the original game with the level k model $50*2/3^k$. Furthermore, an interesting order effect occurs, when a series of p*average games with different known p-values (e.g. 2/3, 1/3, or even negative values like -1/2 and -2/3) are played after a series of noisy beauty contest games without feedback: the resulting average is then close to zero, as if subjects played a game with a public announcement with $\varepsilon^i=0$ for all players. Thus, Benhabib et al. (2016) solve Keynes' level k "problem" by introducing Keynesian sentiments, formulated as ε^i, to offset the need of subjects to do complicated calculations. The mathematical proofs, however, are more complicated than in the original beauty contest game.

Yet, anchors can also be misleading: if e.g., ε^i is drawn i.i.d. from $\mathcal{N}(c, \sigma^2)$, c>0, then the equilibrium is $(2/3*c/(1-2/3))+\varepsilon^i$, (see Benhabib et al., 2016). In this still untested game, many subjects would probably also choose (near) their signals, as a simple heuristic.

Notably, when all players receive the same commonly known signal c, c>0, with a target 2/3*average+c, behavior in the first period is not even close to the equilibrium choice (c/(1-2/3)) (see Güth, et al., 2002). Indeed, the anchor c is chosen by many initially, but choices converge towards equilibrium over time. This convergence can be disturbed by changing the constant c to a non-stationary random walk ($c_t = c_{t-1} + \varepsilon_t$ where ε_t is uniformly taken from a known interval). Lambsdorff et al. (2013) implement a price setting game with the target price of 4/5*(average + 5) + $c_t/10$ where 5 is labeled to the subjects as the cost of a raw material and c_t as a "business indicator". Subjects choose the business indicator as a simple heuristic instead of the equilibrium ($20+c_t/2$). In another macroeconomic application, Giamattei and Lambsdorff (2015) use the beauty contest with a known constant as a model of the Keynesian multiplier where c_t is a non-stationary series of exogenous investments. Subjects have to choose individually optimal consumption as 4/5*income, where 4/5 is the marginal propensity to consume. Income is determined as the average consumption + the investment c_t. They find limited reasoning as a driving force of persistent under-consumption.

When c is not common knowledge (interpreted as the state of the economy or the value of a share), but given through a public and private signal (Morris and Shin, 2002), Shapiro et al. (2014) as well as Baeriswyl and Cornand (2014) find that level k is a good descriptive model in standard settings when information is symmetrically distributed among players. However, subjects´ behavior in the asymmetric case is mainly driven by focal points as choosing halfway between the public and the private signal, or choosing one of these two signals. Thus, whether the equilibrium is learnable instantaneously depends very much upon whether the naive players receive an equilibrium strategy as a reference point or not.

An important question or criticism for macro experiments are related to the number of subjects in a setup. Do we need 1000s of players? Combining the player number issue with two different strategic environments, games with strategic substitutes vs complements, Nobuyuki et al. (2016, see also references therein) use further variations of the 2/3*average+c game to analyze convergence to equilibrium. The two games have targets 100-2/3*average (strategic substitute) and 20+2/3*average (strategic complement), respectively, with integer choices from [0,100] and numbers of players varying from n = 2, 3, 4, 5, 8, and 16 with an equilibrium at 60. Logit level k (LLK) and logit cognitive-hierarchy (LCH) model (both allowing for errors in beliefs) predict that deviations of the average choices from the Nash equilibrium is larger in games with strategic complements than with substitutes for n>2. Actual behavior confirms this, except that n>4 and not n>2. Thus, it should not matter for convergence in forecasting experiments (which are similar to those games) whether n is large but rather, that n should not be too small. What matters is the strategic environment.

The main difference between macro models and the micro version of the beauty contest game is that agents' choices contain dynamic elements: a choice in period t may depend not only on the average choice of all players in period t, but also on the beliefs in period t of average choices arising in period t+1. Furthermore, choices are labeled in macro terms, as prices and the average as the inflation rate. For policy making it can be important to figure out good information transmissions and strategic environments, in order to guide behavior quickly to an equilibrium as policies might work well in equilibrium but less so out of equilibrium (Benhabib et al., 2016; see also Woodford, 2013, and García-Schmidt and Woodford, 2015).

The reduced-form New-Keynesian model, which required expectations of future inflation and the output gap and has a central bank setting interest rates under a sticky price assumption, can be reduced to a beauty contest game, since aggregate (average) outcomes can be interpreted as a fraction of the beliefs of future outcomes plus a constant, i.e. an individual guess of inflation today = c + b*average inflation expectation for tomorrow + d*average output gap expectation for tomorrow (see also Duffy, 2016).

4.1.5. Neuro economic and psychology related experiments

Recent work, inspired by the new emerging area of neuroeconomics, raised questions, such as whether different types of reasoning and thus deviations from rational play correlates with certain brain activities (see e.g. Bhatt and Camerer, 2005, in normal form games). Coricelli and Nagel (2009) with BC experiments show that level 1 reasoning produces very different brain activity in some areas as compared to higher levels, as the latter implies a theory of mind (with the medial prefrontal cortex (mPFC) as an important area of the brain), the reasoning about others' reflections while level 1 reasoning assumes no model of the other player, besides random play (being represented in the anterior cingulate cortex (ACC)).

This kind of technique is especially useful, if choices and level of reasoning cannot be clearly separated as in coordination games à la Keynes. Higher activity in mPFC then might suggest some correlation with higher order reasoning (see Nagel et al., 2015, studying choices in lotteries, stag hunt games, entry games). In stag hunt games neither game theoretic reasoning behavior nor brain activity suggest gaining higher profits from a higher level k, all k≥1 choose the same strategy, which is in stark contrast to entry games. In the latter, there is a positive correlation between strategic IQ (level 2 or higher reasoning) related to mPFC activity and payoffs. Also when subjects have the opportunity to learn over time, one can distinguish brain activity between those who use reinforcing strategies vs higher order learning strategies (Hampton et al., 2008), as in E.A Poe's asymmetric intelligence story: "The simpleton had them even upon the first trial, and his amount of cunning is just sufficient to make him have them odd upon the second, I will therefore guess odd; -he guesses odd, and wins" (see appendix IV).

Another new stream of literature includes cognitive measures to understand differences in reasoning. Training in certain cognitive realms like chess players (Bühren and Frank, 2012) shows no difference, while economists obviously play closer to equilibrium (Bosch-Domènech et al., 2002). Dickinson and McElroy (2010) show that sleep deprivation results in diverting even more from equilibrium play; Gill and Prowse (2014) find that subjects with a higher cognitive ability, measured with the Raven test, choose numbers closer to the equilibrium. Further, cognitive ability is improving reasoning in the beauty contest game (Burnham et al., 2009, Brañas-Garza et al., 2012). A single subject beauty contest[13] has been invented by Bosch-Rosa et al. (2015); together with other cognitive measures and the results in the original beauty contest game the authors construct an index to discriminate between "high sophistication" and "low sophistication" subjects. High types amongst themselves (without knowing this grouping) create no bubbles in subsequent in asset markets, unlike when only low types interact with each other.

4.1.6. Leaving the lab and external validity

Since this game is easy to explain, field experiments are abundant, either played at parties, even in Hollywood (Robinson, 2004), in the newspapers (e.g., Thaler, 1997, in *Financial Times*, Bosch and Nagel, 1997, in *Expansion*, Selten and Nagel, 1997, in *Spektrum der Wissenschaft*, which equals the French *Pour la Science* and the *Scientific American;* see also Bosch-Domènech et al., 2002). These papers provide external validity of level k in these field experiments.

[13] In Bosch-Rosa et al. (2015) on the "one-person beauty contest", an unusually high share of subjects (more than 50% of about 350 subjects) find the Nash equilibrium (resp. the payoff maximizing answer). Each subject has to choose two numbers between 0 and 100 and every number is paid according to the distance to 2/3 of the average of the two chosen numbers (cf. our two-person distance treatment in section 5).

In section 2 we mentioned some parallelism between Moulin's class room experiments and Paclet's (1983) analysis of more sophisticated players. Paclet (1983) outlines many reasoning processes, but overlooks a very essential one. The highest probability to be the winner is to run a meta-experiment, best within a similar population you expect for the (newspaper) participants. For example, prior to submitting his number, one such participant from *Spectrum der Wissenschaft* did an online experiment (with 150 subjects) and then submitted a number which was just 0.1 away from the newspaper experiment (with 2,728 subjects). Is that not the best Bayesian type? If you do not have a Bayesian prior, instead of using a naïve one, inventing one, or assuming you know the true distribution, do an experiment with the right representative population, using the resulting distribution as your prior!

There is one line of literature that uses level k models for empirical studies. Natural experiments with entry games induced through Telecommunications Act of 1996 in USA inspired Goldfarb and Xiao (2011) to formulate a structural model of managers' strategic ability based on the cognitive hierarchy model (Camerer and Ho, 2004). They find that higher-ability managers enter into markets with fewer competitors. Goldfarb et al. (2012) and Crawford et al (2013) summarize similar field evidence related to level k and cognitive hierarchy models.

The greatest influence of the game so far is its use in introductory micro (not yet macro) economic theory classes. Also in talks or articles to an (non-economic) audience it appears as a demonstration game to clarify the relation between game theoretic and bounded rational reasoning (see Varian, 2002, relating it to the movie "A Beautiful Mind", in the Ted Talk by Colin Camerer, 2013) or in natural science journals as in Camerer (2003a). The game has made its way into a comic strip (see Coricelli, Nagel, 2010, and Roth, 2009) and forms an entire chapter in a book by Thaler (2015) directed to the general public. Lastly, Levine (2012) finally has come around to accept that level k models are not doomed.

4.2. Related theoretical literature[14]

Since the early 1990s there were several approaches of limited reasoning in micro and macro theory. Recently, level k formulated through experiments entered as a modeling tool into both those areas. Most of the (young) micro theorists in turn test or plan to test their newly developed theories with new experiments.

4.2.1. Bounded rational modeling in micro and (epistemic) game theory

Herbert Simon (1996, p. 34ff) describes that in a Cournot equilibrium tatonnement process "each firm, with limited cleverness, formed an expectation of its competitor's reaction to its actions, but that each

[14] For excellent surveys of related bounded rational models see Crawford et al. (2013) and Crawford (2013). Nagel (1995) cites also Binmore (1987), Auman (1992), Bacharach (1992), and Bicchieri (1993) for similar theoretical limited reasoning models.

carried the analysis only one move deep. But what if one of the firms, or both, tries to take into account the reactions to the reactions? They may be led into an infinite regress of outguessing" Simon (1996, p. 37). Unfortunately, Simon did not know about the beauty contest experiments.[15]

The closest model to level k is Stahl (1993), in which he introduces different levels of smartness of players for two-person games. Smart-0 players choose non-rationalizable strategies or any strategy at random as in the zero intelligence model by Gode and Sunder (1993)[16]. The main idea is that smart-n players know the true population proportion and choices of smart-k agents for k<n and best-respond to them. However, smart-n players have a rationalizability-based model for the behavior of smart-k players for k>n and other k=n agents. They simply know the joint population proportion of all smart-k players for k>n, hold a belief on what those players do, and best-respond to that belief. However, the belief is disciplined: Smart-n players know that smart-k players with k>n will only play strategies which survive at least n-1 steps of what is essentially a rationalizability iterative procedure (Bernheim, 1984, Pearce, 1984). Stahl also provides an evolutionary approach for behavior over time.

Detailing the features of this model exemplifies the indeterminacy a theorist faces when leaving the box of rationality, especially specifying á priori beliefs about type distributions (see also Paclet, 1983, for a similar discussion of the magazine data). As said above, the simple level k model might have disciplined some features of new level k related models, especially that level k players do not specify what higher order players do.

Milgrom and Roberts (1991) discuss learning of sophisticated players starting out with the Cournot best response reasoning; Selten (1991), in his anticipatory learning model also assumes higher-order beliefs, breaking off after some few steps of reasoning about the others players reasoning, which he calls a “speculative attempt, as it is not based on experimental data. This paper inspired Mitzkewitz and Nagel (1993) and also Nagel (1995) to construct anticipation types.

The level k model is now also appearing in epistemic and applied game theory. Epistemic game theory models the beliefs of players explicitly, and considers the behavioral implications of certain restrictions on players' beliefs. For example, a player who reasons up to some finite level may think that her opponents are rational, think that her opponents think that, and so on. Modeling players' beliefs explicitly makes it possible to formalize the idea that, if a game has a solution that is somehow non-obvious or non-focal or as Moulin (1979) states "sophisticated", then players who are bounded in their reasoning may be unable to reason their way toward the solution, even if "unbounded" reasoners can do so (Kets, 2012). It also makes it possible to check whether standard game-theoretic results are robust to relaxing the assumption that players are perfect Bayesian reasoners (Strzalecki, 2014, Heifetz and Kets, 2013).

[15] Nagel (1995), Knoepfle et al. (2009) with eye-tracking technique, and Tang (2001) discuss such anticipation strategies over time.

[16] This zero-intelligence model was based on experimental data in double oral auctions. Herbert Simon actually recommended Shyam Sunder to put a bit of intelligence into his model, which Shyam did not do as his study was about the simplest reasoning. Maybe the level k reasoning follows this recommendation.

Alaoui and Penta (2015) endogenize players' depth of reasoning as stemming from a cost–benefit analysis, and disentangle the player's depth from his beliefs over his opponent's cognitive bound. They also provide axiomatic foundations for the decision to stop reasoning to be represented by a cost–benefit analysis and call it rational unawareness. They test it in experiments described in Alaoui and Penta (2016).

In applied micro Crawford et al. (2009) introduce level k reasoning to the design of optimal auctions, Crawford (2013) for efficient bargaining mechanisms, and Kets and Sandroni (2015) extend the level k model by modeling how a player's identity affects her reasoning, and use this to study the optimal composition of teams.

4.2.2. Level k related models in macro and financial economics

The macro and financial economics literature had referred to the Keynesian beauty contest games, the average opinion game with multiple equilibria long before the experimental literature.[17] We will point to a few papers, but especially refer the interested reader to a recent survey by Angeletos and Lian (2016) who discuss recent macro theory literature on expectation formation and coordination related to beauty contests amongst others, which should provide experimenters a rich set of models that might be worth being tested.

Fryman (1982) discusses Keynes beauty contest games and formulated on page 656 a p-beauty contest game within a supply and demand model and expectation formation of firms. However, since the parameters of the model are not known he can only discuss adaptive learning without specifying higher order beliefs. He refers to Lucas' (1972) parameter specification which allows only first order beliefs. Simonsen (1988) motivates coordination problems in wage settings and consequences for inflation inertia with the 1/2average game.

Many researchers followed the path searching for conditions under which beauty contest equilibria are unique. For example, in games with strategic complementarities (e.g. stag hunt games, which can be considered as a discrete choice beauty contest game with p=1), Carlson and Van Damme, (1993) and Morris and Shin (1998) introduce incomplete information about payoffs and subjects receive noisy signals about these payoffs to obtain uniqueness. Acemoglu and Jensen (2010) study interesting varieties of beauty contest games with different types of noise structures in strategic complement and substitute environments, and provide comparative statics and equilibrium properties.

Woodford (2003) introduced the idea of higher order beliefs in games with strategic complementarities and provides ample discussions for policy implications. Allen et al. (2006) derive different orders of

[17] Most macro theorists do not yet know the experimental (beauty contest) literature, even if the experiments build on related macro theories.

beliefs in a finance context of asset market speculation with risk-averse and short-lived traders. They build on Tirole (1982), who shows that the existence of short-term traders may explain why price bubbles occur in asset markets. Allen et al. (2006) state: "Do asset prices reflect average opinion, and average opinion about average opinion, in the manner that Keynes suggests? Our answer is a resounding ‘yes’." This has implications for the mean path of prices which will deviate from the consensus expectations of fundamentals and therefore have some elements of a bubble, and also prices exhibit inertia in reacting to changes in the fundamental value of the asset. Nimark (2011, see also references therein on higher order belief modeling) proves that already a finite number of higher order beliefs (around 4 to 10 steps) are typically close to an equilibrium of infinite regress in asset markets. However, he assumes that human players behave as if they are able to form such believes which our experimental results largely reject.

However, already in the recent few years the gap between macro bounded rational modeling and experimental evidence is starting to close. There are two boundedly rational reasoning and learning concepts that show some similarity but contain important differences with respect to level k. The first has been developed by Evans and Ramey (1992)[18] who propose that agents use a finite number of iterations of the mapping between beliefs and outcomes, but the level of reasoning is, in contrast to level k, not determined by strategic considerations, but endogenously by calculation costs. The second approach, frequently referred to as "eductive learning", both in the micro and macro literature, has been proposed by Guesnerie (1992), who investigates whether rational expectations equilibria are learnable by the "eductive learning", assuming that individuals constantly adjust their levels of reasoning, which has, however, no stopping rule. Guesnerie (2008) analyzes the “eductive stability” of perfect foresight equilibrium in the context of monetary policy rules.

Woodford (2013) reviews and further investigates Evans and Ramey’s and Guesnerie’s approaches in a neo-Keynesian framework, and shows that the speed of convergence depends on the nature of the monetary policy rule. In order to reconcile the missing increase in inflation after the financial crisis after 2007 with New-Keynesian economics, García-Schmidt and Woodford (2015), referring amongst others to the experimental papers on level k, expand on a continuous level k model (which could be interpreted as an average of several discrete level k types) and study learning dynamics. Most importantly the possibility of slow convergence is not analogous to Guesnerie's argument for the failure of eductive stability. Instead they find slow convergence for the policy rules that Guesnerie calls "eductively stable", but faster convergence for the ones (with a stronger inflation-response coefficient) that he finds to be "unstable”. There is at least one inconsistency worth mentioning with respect to their bounded rational model: While García-Schmidt and Woodford (2015) assume that agents apply a finite number of iterations in the strategic dimension when thinking about other players’ reasoning, they maintain the

[18] Evans in a private communication stated that he did not continue with his level k idea since he could not link it with empirical evidence; instead the learning model in the same paper became an important modeling tool for macroeconomists interested in bounded rational learning.

assumption, because of mathematical difficulty otherwise, that agents maximize utility and profits over an infinite horizon so that even agents with low levels of reasoning still form an infinite sequence of beliefs about market outcomes at any future date.

In related subsequent work, Farhi and Werning (2016) introduce level k into an overlapping generation model, with occasionally binding liquidity constraints to circumvent the reasoning by agents about the infinite horizon typically faced in these models. They show that monetary policy (specifically, commitments about future interest-rate policy, i.e., forward guidance) has particularly weak effects when level k is combined with incomplete markets while rational expectation analysis would predict that it has a rather strong effect. These disagreements of convergence between different theoretical modeling techniques, now incorporating bounded rational reasoning, clearly call for careful empirical validation.

In this section we have shown several indeterminacies when it comes to making predictions (about real world economic situations): 1. Which model specifications to choose. 2. Which (bounded) rational reasoning procedure to choose. 3. The outcome can be different for different procedures within the same model.

In sum, Moulin's textbook was the first step in the development of this stream of beauty contest research guiding the experimental economics literature. Step 2 was an experimental design that produced few observations of equilibrium play. Step 3 was research that is not only motivated by challenging game theory's power to predict behavior, but by a wish to understand what else explains subjects' behavior. Finally, the experimental results and behavioral models developed in the past 20 years have established behavioral foundations for new models in micro and macro theory which ask for new empirical validations. Macro and micro theorists and experimenters are beginning to complement each other's work.

Given the main theme of this paper, the experiment described in the next section refers to Ledoux (1981) and Moulin (1986). Unlike suggested by them, it makes surprisingly little behavioral difference whether players have a dominant equilibrium strategy or not, or whether they play a 2 person or many person beauty contest variant. We show that level k is the best descriptive model, although three different treatments produce three different logical reasoning procedures. This sheds further light on subjects' perceptions of the rules of the game.

5. A new two-person beauty contest experiment

Ledoux (1981) mentioned in passing that zero is a reasonable choice in a two-person game with one winner. Moulin (1986, p. 74f) comments: "it [finding the sophisticated equilibrium] presupposes that each and every player either computes all dominated strategies of all players... [which] does not cause much trouble." (see also his quote in the introduction, and section 2).

By contrast, experimental results show that many subjects, students and economists, turned out to play (weakly) dominated strategies (Grosskopf and Nagel, 2008) in 2 person fixed payoff treatments. Chou et al. (2009) interpret this finding as showing that game form recognition might be an important concept to understand out of equilibrium behavior. If the game is not understood or another game than the experimenter proposes is played, then one cannot expect the players to adopt a rational choice.

With a new little twist, we show one possible reason for the emergence of out-of-equilibrium behavior with level k as the unifying approach: What happens if subjects misunderstood the payoff structure or played a game having an entirely different payoff structure in mind, not paying attention to other features of the game?

5.1 Experimental design and equilibrium

We use the two-person design from Grosskopf and Nagel (2008) with a fixed price payoff, and then implement the payoff structure from Benhabib et al. (2016) with a continuous payoff structure.[19] Two players have to choose a real number in the closed interval of [0,100], with the highest payoff to the person being closest to 2/3 times the average choice.[20] Additionally, we compare our data with the data of the different treatments with n>2 from Bosch -Domènech et al. (2002), Bühren and Frank (2012), and with professionals with n=2 in Grosskopf and Nagel (2008).

2 person treatments:

Fixed payoff: Only the person closest to the target, 2/3 times the average of both chosen numbers, receives a fixed price of 10 Euros, with the prize split if tied (Grosskopf and Nagel, 2008).

Distance payoff: Both players are paid according to their distance to 2/3 times the average of both numbers with the payoffs π(i) of player i (i=1,2) being

$\pi(\mathrm{i}) = 100 - \left(x(i) - \frac{2}{3}\frac{x(1)+x(2)}{2}\right)^2$, where x(.) is the choice of a player, 100 points being 5 Euros.

[19] Güth et al. (2002) also used a distance payoff and compared it with the tournament payoff structure for n>2 players without finding differences in behavior. However, they do not discuss the difference in the weakly vs. strictly dominated strategies or the difference in the Pareto optimality conditions.

[20] Costa-Gomes and Crawford (2006) play a two-person game with distant payoffs in which each player has his own p parameter and intervals of choices. The player has to guess p*other choice. They carefully analyse the reasoning processes arising through plays in 16 different games without feedback between games. Instead of making the parameters common knowledge, subjects need to inquire them by mouse clicks on boxes. This way the different reasoning procedures can be more clearly analysed.

Equilibrium:

Fixed payoff: Both players choosing zero is a unique equilibrium in weakly dominant strategies. To see this, consider that the two players choose a and b respectively, a<b. Since 2/3 of the average will be $\frac{1}{3}(a+b)$, the game is isomorphic to the game "the lower number always wins". Any strategy combination is Pareto optimal.

Distance payoff: Both players choosing zero is a unique Pareto optimal equilibrium. Out of equilibrium it is optimal to play a=b/2, since a=2/3*(a+b)/2, when a<b. Zero is the only rationalizable strategy obtained through an iterated elimination of *strictly dominant* strategies, as in the n>2 case, however, the different iteration steps are now 100, 50, 25 etc.[21]; or starting from focal point 50, it is 25, 12.5 etc.

The continuous payoff structure with its property of iterated eliminating *strictly* dominated strategies (IESDS), requires weaker (or more basic) epistemic conditions than the iterated elimination of weakly dominated strategies. IESDS captures the implications of rationality and common belief in rationality, so rationality and common belief of rationality by itself does not imply equilibrium play in the standard beauty contest game, unlike in the game where the equilibrium outcome can be obtained by IESDS. This two-person game could have appeared in Moulin's sophisticated equilibrium chapter unlike the seemingly simple fixed payoff two-person game with a (weakly) dominant strategy equilibrium.

To summarize the logical reasoning procedures for the three (or four) different treatments:

1. for n=2, fixed payoff Players have a (weakly) dominant strategy zero.

2. for n=2, distant payoff: The equilibrium is reached through a procedure of iterated elimination of strictly dominated strategies with multiplication parameter ½, starting at 100.

3. (and 4.) for n>2, fixed payoff, (distant payoff): The equilibrium is reached through a procedure of iterated elimination of weakly (strictly) dominated strategies starting at 100.

During a lecture on introduction to economics, we conducted a one-period version in Kassel in February 2016 with 309 undergraduate students (160 with fixed and 149 with distance payoff). The experiment was programmed with qualtrics.com, an online program typically used for surveys, and students accessed it via their smartphones. The allocation to treatments and the matching of partners was randomized. Four student helpers made sure that our subjects did not talk to each other. We paid 20 couples in fixed and 20 couples in distance payoff: In the fixed payoff treatment, we paid 10 Euros to

[21] Note in the case of n>2 and fixed payoff the equilibrium is reached through iterated elimination of *weakly* dominant strategies.

the winner (or 5 Euros to both if they picked the same number). In the second treatment, we paid both players 5 Euros if their numbers were exactly 2/3 of the average and less if the number of a subject was farther away from 2/3 of the average (see distance payoff formula above). Note that in equilibrium the payoffs of both treatments are identical.

5.2. Experimental result

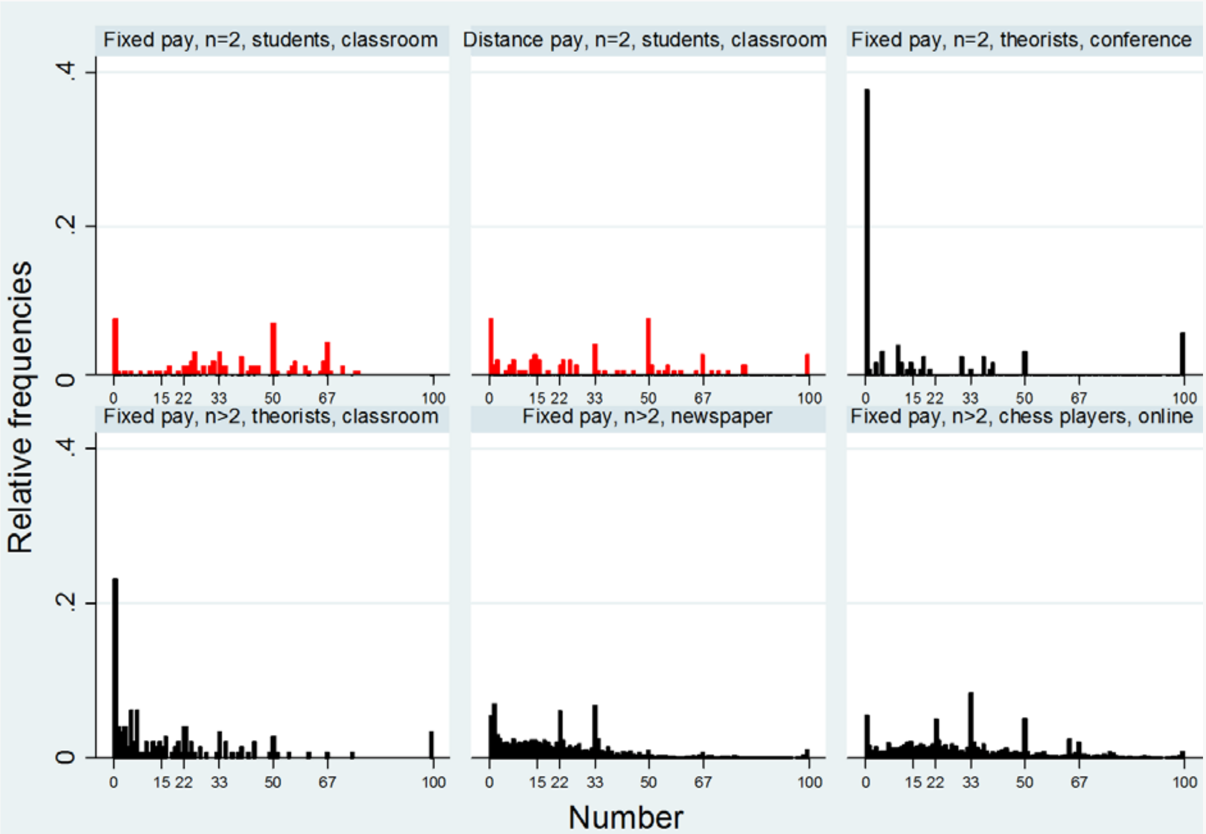


Figure 1: Relative frequencies of choices in the two-person games with fixed payoffs (upper left), distance payoffs (upper middle) and with economics professors in various conferences (upper right); lower left n>2 games with economic professors and students in advanced economic classes; in lower middle newspaper contestants; lower right chess players online. Data source: UR: Grosskopf, Nagel (2008); LL and LM: Bosch-Domènech et al. (2002); LR: Bühren and Frank (2012)

The relative frequencies resulting from the three treatments (including n>2 treatments), distinguishing also different subject pools, are shown in Figure 1: The upper graphs show different n=2 treatments and the lower graphs n>2 treatments. All figures feature the same spikes, at or near 67, 50, 33, 22, and 0, thus the same reasoning procedure described by the same level k model. This is in stark contrast to the 4 different logical reasoning procedures we described above. This means that in 2 person games, most subjects cannot find the dominant strategy zero when applicable and also are not aware of their own influence on the average choice through their own number.

For n=2, the average number in fixed payoff is 36.98 (std. dev. 22.72) and in distance payoff 37.36 (std. dev. 28.04)[22], with no significant difference according to a two-sided t-test and a two-sided Mann-Whitney U test, and a two-sample Kolmogorov-Smirnov test. Further, the shares of choosing the Nash equilibrium are not significantly different across treatments according to a Fisher exact test. Figure 1 illustrates also the choices of theorists (professional sample) in the two-person beauty contests of Grosskopf and Nagel (2008) as well as theorists of the n>2 beauty contests of Bosch-Domènech et al. (2002) (upper right and lower left histograms). Furthermore, it shows the entries of the newspaper beauty contests of Bosch-Domènech et al. (2002) and of the online beauty contest with chess players of Bühren and Frank (2012) (lower middle and lower right histograms).

Comparing our average numbers to those in the student samples of the two-person guessing game of Grosskopf and Nagel (2008) with fixed payoffs (student sample) and the lab experiments (n>2) reported in Bosch-Domènech et al. (2002), two-sided t-tests detect no significant differences. Yet the average number in the professional sample of Grosskopf and Nagel (2008), those of the classroom, take-home, theorists, internet, and newspaper experiments reported in Bosch-Domènech et al. (2002) as well as the average number stated by chess players in the online experiment of Bühren and Frank (2012) are significantly lower than the averages in our two treatments analyzed above. Lower numbers in take-home, online, and newspaper experiments can be explained by response time: Rubinstein (2007) showed that more reasonable guesses in the beauty contest are derived after larger response times.

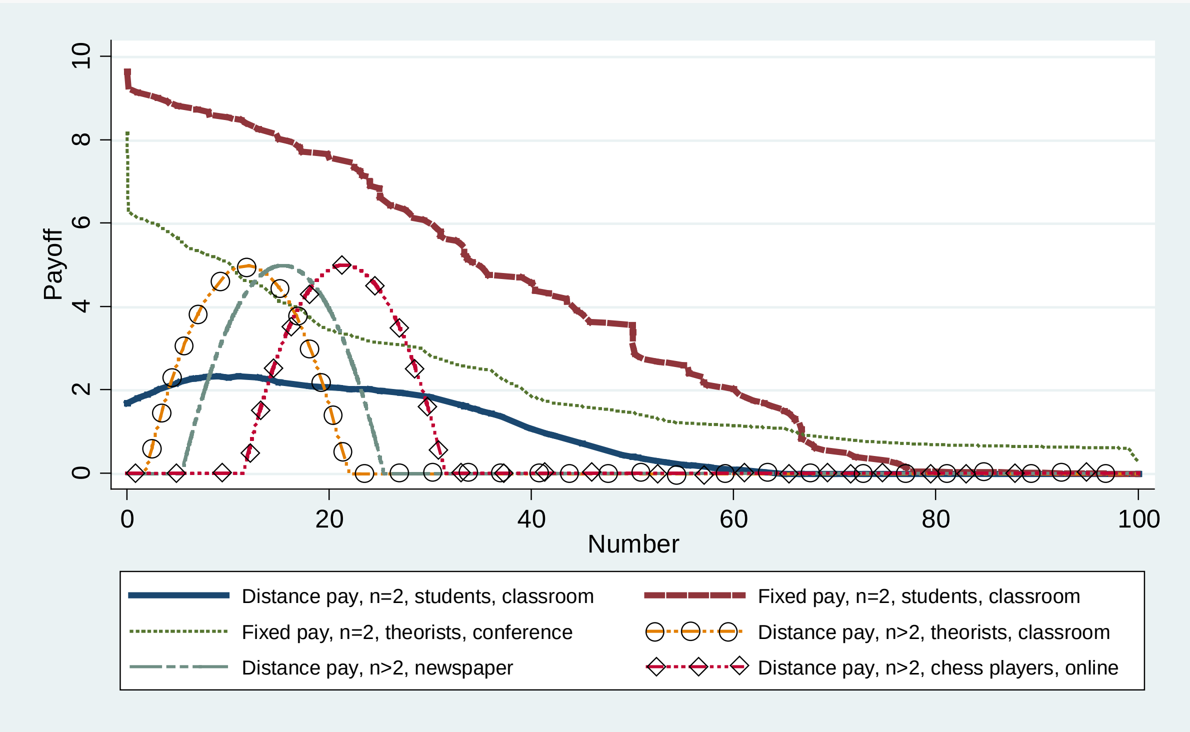


Figure 2: Distribution of payoffs by number for the experiments of Figure 1

[22] The raw data and the subjects' comments of our two new treatments can be found online in the supplementary material of this Nagel et al. (2016).

Figure 2 compares possible payoffs of stated numbers in each of the experiments of Figure 1: We used a recombination method for the two-person beauty contests by recombining all possible matching couples. Then we calculated the expected payoff for each actual choice according to the fixed and distance payoff scheme, respectively. For n>2 beauty contests, we did not apply the recombination method. In order to better compare these graphs to n=2 graphs, we calculated the payoffs according to our distance payoff function (with 2/3 of the average of all given numbers, see above) for each single session and averaged across all sessions within the same treatment.[23]

Figure 2 depicts inverted U shaped graphs for n>2 beauty contest games with a peak around 2/3 of the average for the three populations with n>2. The graph of our two-person distance treatment is much flatter with a peak at number 11.05 (2/3 of the average was 24.91). The graphs of the two-person beauty contests with fixed payoffs are monotonic decreasing – in the theorist sample with lower payoffs for small numbers than in our student sample because a lot of theorists at conferences chose the Nash equilibrium zero (see also Figure 1). Thus they had to split their prize when matched with another zero and often lost with small numbers (when matched with a zero).

Our little experiment shows the fundamental problem of game theory when it comes to making predictions at least in the first period(s).[24] Small changes in the parameters of the game can result in significant differences in the properties of the game, yet unperceived by the human players who might use an entirely different reasoning mechanism compared to the theoretical arguments. On the other hand, technical changes which leave the game-theoretic properties unchanged might result in considerable changes in human behavior. For instance, behavior differs a lot if a prisoners' dilemma is labeled "Community Game" or "Wall Street Game" (see e.g. Liberman et al., 2004, and Levitt and List, 2007, for an overview). However, game theory can suggest optimal behavior, especially when there is a dominant strategy as in the two-person fixed payoff game. Furthermore, it suggests how to structure the strategy space, as in the case of iterated reasoning, which generates useful benchmark structures.

The main reason for the deviation from equilibrium is that players start out with focal points far from equilibrium and apply low level of reasoning. Benhabib et al. (2016) delete the boundaries 0 and 100 to make zero a focal point, by allowing all numbers, positive, negative and zero, with the same target of the original beauty contest game. The median shrinks to less than 15 in a one shot game while the typical median in the original game in undergraduate classes is around 33. The equilibrium is the same, but it is not reached by an iterated elimination process as there is no upper limit which eliminates at least the focal point, far away from equilibrium. Clever idiosyncratic signals as in Benhabib et al. (2016) as mentioned in section 4. 1. 4. turn the signals into optimal anchors leading to high payoffs for all. Also

[23] Güth et al. (2002) showed that beauty contest distributions of choices with distance and fixed payoffs are not significantly different (see footnote 16 above).
[24] Grosskopf and Nagel (2009) show slow convergence over time also in the two-person guessing game with fixed payoff.

when the multiplication parameter is negative, e.g., -2/3 instead of 2/3, behavior is closer to equilibrium with averages near -5, as it is a game of strategic substitute (see discussion in 4.1.4.).

An application and a word of warning regarding these findings should be considered for example in mechanism design issues, where using a dominant strategy seems like a preferred simple choice to implement the mechanism. In the light of our simple experiment, the reader is referred also to Camerer (2003a) and his discussion on Glazer and Rosenthal (1992) and Abreu and Matsushima (1992) with their divergent ideas about the implementability of the Abreu–Matsushima mechanism. Amongst others, it is argued that only in the simplest case with an equilibrium in dominant strategies the mechanism might work. However, in the light of our results, it could be that bounded rational agents might not even find this obvious and reasonable strategy, as in our case (see also the discussion by Crawford et al. 2009). Instead the theorist needs to also consider obvious focal points which deflect even clever agents' reactions, as these might give best response to low level reasoners, although a dominance argument is present.

6. Conclusion

In his blog, Alvin Roth (2009), referring to our discovery of the beauty contest's invention, quotes from Roth et al. (1990, p. 170): "What is important about Columbus' discovery of America is not that it was the first, but that it was the last. After Columbus, America was never lost again." The beauty of this analogy comes at the price of imprecision. It would match our story better had the Vikings published an account of their discovery in some popular yet remote Scandinavian geography journal, which an alert scholar was curious enough to read, later passing it on in his geography textbook in terms the Vikings never heard of. In any case, Alain Ledoux was right when he wrote to the experimenters (our English translation): "I have been very amused to respond to your question suggested at the ChessBase site, which we (Bühren and Frank, 2012) used to conduct the experiment. For I think well to be … the inventor", and he should be given credit for initiating research around the guessing game.

Yet it was Hervé Moulin who discovered the potential of the game and transmitted it to the profession in game theoretic language. He inspired lecturers in economics, who in turn inspired Rosemarie Nagel. Nagel (1993, 1995) turned Moulin´s "guess the average" thought experiment arguably into a proper lab experiment and provided through actual data one of the fundaments of bounded rational reasoning: limited reasoning as described by Keynes (1936). Yet, we also showed that macroeconomists were directly inspired by Keynes's metaphor and took a somewhat different turn to introduce it into economics. Finally these two approaches seem now converging, hopefully, more in Keynes's spirit.

Nagel's personal account of being a subject and formulating a descriptive model as a result of the experience might suggest consequences for higher education in economics. In most other social and natural sciences, even undergraduates get involved in experimenting, either as subjects or in laboratory tutorial. Economics students also should get this chance before they enter the state of curse of knowledge about theoretical outcomes. Even experimenters cannot always make sense of subjects' behavior, observed from the outside, but very easy to discover as a subject.

The main advantage of the beauty contest game in comparison to earlier experiments on normal form games with similar structures was that the former contains a continuous strategy space with a clearer differentiation between different choices emerging from different level k's within the same game and no conflicting other regarding preferences. Also most games studied with dominance structures (in the early nineties) had too few strategies to go beyond level 3 reasoning. However, new experiments, as e.g. the 11 to 20 game, offset these difficulties. Given that key features in economic activity are driven by the beliefs and (aggregate) actions of agents, the beauty contest game still has a great potential for uncovering strategic biases in many different situations, especially if they are embedded in new and larger environments as in several models (especially in macro contexts) discussed in this paper.

The most important critical ingredient became the level 0 reasoning, which can be related to random choice, intuitive answers, as the signal in auctions, etc., maybe evoked by an automatic reasoning procedure. This seems arbitrary and thus has rightly been a major criticism of this theory. However, it is the art of the reasoner to figure out what the simpleton will do in many kinds of situations and whether this simpleton will survive over time or in one's own reflections. If level 0 is constructed ex post, given a data set, then later experiments have to demonstrate the robustness of such findings. We think Keynes's beauty contest makes precisely this point of figuring out the best reply to a multiplicity of thoughts in an iterated way. After Schelling (1960), any further insights have to be accompanied through experiments in the lab and in the field but not in the armchair (Simon and Bartel, 1986). Maybe then we will find a general theory of simple heuristics for level 0 thinking.

The descriptive level k models have been one of main breakthrough to structure behavior in many different games, especially when behavior is not in equilibrium. Our results on two different two-person games with fixed vs distance payoff compared with the n>2 games provide an example for how difficult it can be to find a dominant strategy. Pareto inefficiency of the equilibrium is not the hindrance, as it is in the prisoner's dilemma supergames argument. Behavior is truly bounded rational. Our three different treatments with three different mathematical reasoning concepts can be described with one descriptive model, the level k model.

Furthermore, new theories have been formulated in epistemic game theory, applied micro and macro based on these findings. In the early 90ties, many theoretical papers used iterated reasoning. However, just as in games with multiplicity, theorists could not solve the numerous arising indeterminacies when to stop the process for bounded rational reasoning. In the macro and finance literature, there are many games which imbed beauty contest features, largely unstudied by experimental economists (e.g. Angeletos et al., 2010, Cespa and Vives, 2015). Such models can foster new links between theorists with great modeling tools and experimenters with equally useful tools, selecting empirically relevant behaviors in the two cases of indeterminacy, through multiple equilibria or multiple bounded rational reasoning procedures. There is already such a small but growing and fruitful interaction in macroeconomics (see survey by Duffy (2016), and on experiments featuring beauty contest elements in particular (e.g., Benhabib et al., 2016, or in new Keynesian models as in Pfajfer and Zakelj, 2011, Hommes and Zhu, 2014, Mauersberger, 2016).

The rules of guessing games are simple to state, though most difficult to answer. The level k model and its variations are easy to apply, even in everyday situations or on reasonings such as Seneca's conjecture "Nil sapientiae odiosius acumine nimio." (Nothing is more hateful to wisdom than excessive cleverness, quoted in Poe, 1845). We interpret excessive cleverness with excessive iterations and thus that limited level k is an important part of human nature. Connecting our findings with those of the humanities and novelists, who implemented bounded rational reasoning in richer contexts and long before economists have (mathematically) structured it, can be another new adventure.

All hats off to Hervé Moulin, for each time the game is played, studied, and reflected upon one more ray of the truth shines on us. Happy birthday, Hervé Moulin, and thank you for this game.

References

Abreu, D., & Matsushima, H. (1992). Virtual implementation in iteratively undominated strategies: complete information. Econometrica, 60(5), 993-1008.

Acemoglu, D., & Jensen, M. K. (2010). Robust comparative statics in large static games. In 49th IEEE Conference on Decision and Control, December, 3133-3139.

Agranov, M., Caplin, A., & Tergiman, C. (2015). Naive play and the process of choice in guessing games. Journal of the Economic Science Association, 1(2), 146-157.

Alaoui, L., & Penta, A. (2016). Endogenous depth of reasoning. Review of Economic Studies, rdv052, forthcoming. doi: 10.1093/restud/rdv052

Alaoui, L., & Penta, A. (2015). Cost-Benefit Analysis in Reasoning, mimeo.

Allen, F., Morris, S., & Shin, H. S. (2006). Beauty contests and iterated expectations in asset markets. Review of Financial Studies, 19(3), 719-752.

Angeletos, G. M., Lorenzoni, G., & Pavan, A. (2010). Beauty contests and irrational exuberance: A neoclassical approach (No. w15883). National Bureau of Economic Research.

Angeletos, G. M., & Lian, C. (2016). Incomplete Information in Macroeconomics: Accommodating Frictions in Coordination. Prepared for the Handbook of Macroeconomics, Vol. 2. Editors, John Taylor and Harald Uhlig.

Arad, A., & Rubinstein, A. (2012). The 11–20 money request game: a level k reasoning study. The American Economic Review, 102(7), 3561-3573.

Arad, A., & Rubinstein, A. (2012). Multi-dimensional iterative reasoning in action: The case of the Colonel Blotto game. Journal of Economic Behavior & Organization 84 (2012) 571– 585

Arjona, D.R., Fragiadakis, D., & Kovaliukaite, A. (2016): Describing individual beliefs in a strategic environment: Evidence from eliciting subjects´ predictions on how a population plays. Mimeo.

Aumann, R. J. (1992). Irrationality in game theory. Economic analysis of markets and games: Essays in honor of Frank Hahn, 214-27.

Bacharach, M. (1992). Backward induction and beliefs about oneself. Synthese, 91(3), 247-284.

Baeriswyl, R., & Cornand, C. (2014). Reducing overreaction to central banks' disclosures: theory and experiment. Journal of the European Economic Association, 12(4), 1087-1126.

Bayona, A., Brandts, J., & Vives, X. (2016). Supply Function Competition, Private Information, and Market Power: A Laboratory Study.

Benhabib, J. (2016). A Note on Correlated Equilibria in Beauty Contest Games. mimeo.

Benhabib, J., Duffy, J., &Nagel, R. (2016). De-framing rules to de-anchor beliefs in Beauty contest games, mimeo.

Benhabib, J., Wang, P., & Wen, Y. (2015). Sentiments and aggregate demand fluctuations. Econometrica, 83(2), 549-585.

Bergstrom, T. C. (2003). Vernon Smith's Insomnia and the Dawn of Economics as Experimental Science. Scasenecandinavian Journal of Economics 105, 181-205.

Bernheim, B. D. (1984). Rationalizable strategic behavior. Econometrica, 52(4), 1007-1028.

Bhatt, M., & Camerer, C. F. (2005). Self-referential thinking and equilibrium as states of mind in games: fMRI evidence. Games and Economic Behavior, 52(2), 424-459.

Bicchieri, C. (1993). Rationality and coordination. Cambridge University Press.

Binmore, K. (1987). Modeling rational players: Part I. Economics and philosophy, 3(02), 179-214.

Borel, E., 1921. La Theorie du Jeu et les Equations Integrales a Noyau Symetrique. Comptes Rendus Hebdomnadairs des Seances. Academie des Sciences 173, 1304–1308 (translated by L.J. Savage as "The theory of play and integral equations with skew symmetric kernels″ Econometrica 21(1953), 97–100).

Bosch and Nagel (1997). El Juego de Adivinar el Número X: Una Explicación y la Proclamación del Vencedor." Expansion, June 16, pp. 42-43.

Bosch-Domènech, A., Montalvo, J. G., Nagel, R., & Satorra, A. (2002). One, two, (three), infinity,...: Newspaper and lab beauty-contest experiments. American Economic Review, 92(5), 1687-1701.

Bosch-Domènech, A., Montalvo, J. G., Nagel, R., & Satorra, A. (2010). A finite mixture analysis of beauty-contest data using generalized beta distributions. Experimental Economics, 13(4), 461-475.

Bosch-Rosa, C., Meissner, T., & Bosch-Domènech, A. (2015). Cognitive bubbles. http://ssrn.com/abstract=2553230

Brañas-Garza, P., García-Muñoz, T., & González, R. H. (2012). Cognitive effort in the beauty contest game. Journal of Economic Behavior & Organization, 83(2), 254-260.

Breitmoser, Y. (2012). Strategic reasoning in p-beauty contests. Games and Economic Behavior, 75(2), 555-569.

Brocas, I., Carrillo, J. D., Wang, S. W., & Camerer, C. F. (2014). Imperfect choice or imperfect attention? Understanding strategic thinking in private information games. Review of Economic Studies, 81(3), 944-970.

Bühren, C. & Frank, B. (2012). Chess players' performance beyond 64 squares: A case study on the limitations of cognitive abilities transfer, Talent Development and Excellence, 4(2), 157-169.

Burchardi, K. B., & Penczynski, S. P. (2014). Out of your mind: Eliciting individual reasoning in one shot games. Games and Economic Behavior, 84, 39-57.

Burnham, T. C., Cesarini, D., Johannesson, M., Lichtenstein, P., & Wallace, B. (2009). Higher cognitive ability is associated with lower entries in a p-beauty contest. Journal of Economic Behavior and Organization, 72(1), 171-175.

Camerer, C. (2003a). Behavioral game theory: Experiments in strategic interaction. Princeton University Press.

Camerer, C. F. (2003b). Behavioural studies of strategic thinking in games. Trends in Cognitive Sciences, 7(5), 225-231.

Camerer, C. (2013). When you're making a deal, what's going on in your brain?. Ted Talk. https://www.ted.com/talks/colin_camerer_neuroscience_game_theory_monkeys

Camerer, C., & Ho, T.-H. (1998). Experience-weighted attraction learning in coordination games: Probability rules, heterogeneity, and time-variation. Journal of Mathematical Psychology, 42(2), 305-326.

Camerer, C. F., Ho, T. H., & Chong, J. K. (2004). A cognitive hierarchy model of games. Quarterly Journal of Economics, 119(3), 861-898.

Carlsson, H., & Van Damme, E. (1993). Global games and equilibrium selection. Econometrica, 61(5), 989-1018.

Cespa, G, & X. Vives (2015). The Beauty Contest and Short-Term Trading. Journal of Finance, 70(5), pp 2099-2153.

Chou, E., McConnell, M., Nagel, R., & Plott, C. R. (2009), The control of game form recognition in experiments: Understanding dominant strategy failures in a simple two-person "guessing" game. Experimental Economics, 12(2), 159–179.

Colman, A. M., Pulford, B. D., & Lawrence, C. L. (2014). Explaining strategic coordination: Cognitive hierarchy theory, strong Stackelberg reasoning, and team reasoning. Decision, 1(1), 35.

Coricelli, G., & Nagel, R. (2009). Neural correlates of depth of strategic reasoning in medial prefrontal cortex. Proceedings of the National Academy of Sciences, 106(23), 9163-9168.

Coricelli, G., & Nagel, R. (2010). Walking with Reinhard Selten and the Guessing Game: From the Origin to the Brain. In The Selten School of Behavioral Economics (pp. 111-123). Springer Berlin Heidelberg.

Costa-Gomes, M., Crawford, V. P., & Broseta, B. (2001). Cognition and behavior in normal-form games: An experimental study. Econometrica, 69(5), 1193-1235.

Costa-Gomes, M. A., & Crawford, V. P. (2006). Cognition and behavior in two-person guessing games: An experimental study. The American economic review, 96(5), 1737-1768.

Crawford, V. P., & Iriberri, N. (2007). Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions?. Econometrica, 75(6), 1721-1770.

Crawford, V. P., Kugler, T., Neeman, Z. and Pauzner, A. (2009). Behaviorally Optimal Auction Design: Examples and Observations. Journal of the European Economic Association 7 (2–3): 377-387.

Crawford, V. P., Costa-Gomes, M. A., & Iriberri, N. (2013). Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. Journal of Economic Literature, 51(1), 5-62.

Crawford, V. P. (2015). Efficient mechanisms for level k bilateral trading. mimeo: http://econweb.ucsd.edu/~v2crawford/LkBilateralTrading1April15.pdf

Dickinson, D. L., & McElroy, T. (2010). Rationality around the clock: Sleep and time-of-day effects on guessing game responses. Economics Letters, 108(2), 245-248.

Duffy, J. & Nagel, R. (1997). On the robustness of behaviour in experimental ‘beauty contest’ games, Economic Journal, 107(445), 1684-1700.

Duffy, J. (forthcoming 2016). Macroeconomics: A Survey of Laboratory Research. In Handbook of Experimental Economics, Volume 2, J.H. Kagel and A.E. Roth (eds.), Princeton University Press.

Evans, G. W., & Ramey, G. (1992). Expectation calculation and macroeconomic dynamics. American Economic Review, 82(1), 207-224.

Farhi, E., & Werning, G. (2016). The Power of Forward Guidance with Bounded Rationality and Incomplete Markets. mimeo

Falkinger, J. (1988). Kreativität und Sensibilität versus Interdisziplinarität. Jahrbuch für Neue Politische Ökonomie, 7, 3-19.

Fragiadakis, D. E., Knoepfle, D. T., & Niederle, M. (2013). Identifying predictable players: Relating behavioral types and subjects with deterministic rules. mimeo.

Friedenberg, A., Kets, W., & Kneeland, T. (2015). Bounded Reasoning: Rationality or Cognition, mimeo.

Fryman, R. (1982). Towards an Understanding of Market Processes: Individual Expectations, Learning and Convergence to Rational Expectations Equilibrium. American Economic Review 72, pp. 652-88. Reprinted in Recent Developments in Macroeconomics, edited by E.S. Phelps, Edward Elgar Publishing.

García-Schmidt, M., & Woodford, M. (2015). Are low interest rates deflationary? A paradox of perfect-foresight analysis. No. w21614, National Bureau of Economic Research.

Georganas, S., & Nagel, R. (2011). Auctions with toeholds: An experimental study of company takeovers. International Journal of Industrial Organization, 29(1), 34-45.

Giamattei, M., & Lambsdorff, J. Graf (2015): Balancing the Current Account - Experimental Evidence on Underconsumption. Experimental Economics 18(4): 679-696.

Gill, D., & Prowse, V. L. (2014). Cognitive ability, character skills, and learning to play equilibrium: A level k analysis. Journal of Political Economy, forthcoming.

Glazer, J., & Rosenthal, R. W. (1992). A note on Abreu-Matsushima mechanisms. Econometrica, 60(6), 1435-1438.

Gode, D. K., & Sunder, S. (1993). Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality. Journal of Political Economy, 101(1), 119-137.

Goeree, J. K., & Holt, C. A. (2004). A model of noisy introspection. Games and Economic Behavior, 46(2), 365-382.

Goldfarb, A., & Xiao, M. (2011). Who thinks about the competition? Managerial ability and strategic entry in US local telephone markets. The American Economic Review, 101(7), 3130-3161.

Goldfarb, A., Ho, T. H., Amaldoss, W., Brown, A. L., Chen, Y., Cui, T. H., ... & Xiao, M. (2012). Behavioral models of managerial decision-making. Marketing Letters, 23(2), 405-421.

Grosskopf, B., & Nagel, R. (2007). Rational reasoning or adaptive behavior? Evidence from two-person beauty contest games. Evidence from Two-Person Beauty Contest Games (June 2007). Harvard NOM Research Paper, (01-09).

Grosskopf, Brit & Nagel, R. (2008). The Two-Person Beauty Contest, Games and Economic Behavior, 62(1), 93–99.

Guesnerie, R. (1992). An exploration of the eductive justifications of the rational-expectations hypothesis. American Economic Review, 82(5), 1254-1278.

Guesnerie, R. (2008). Macroeconomic and Monetary Policies from the Eductive Viewpoint. Documentos de Trabajo (Banco Central de Chile), (498), 1.

Güth, W., Kocher, M., & Sutter, M. (2002). Experimental ‘beauty contests’ with homogeneous and heterogeneous players and with interior and boundary equilibria. Economics Letters, 74(2), 219-228.

Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2008). Neural correlates of mentalizing-related computations during strategic interactions in humans. Proceedings of the National Academy of Sciences, 105(18), 6741-6746.

Haruvy, E., D. Stahl, and P. Wilson (1999). Evidence for Optimistic and Pessimistic Behavior in Normal-Form Games. Economics Letters 63, 255-259.

Haruvy, E. (2002). Identification and Testing of Modes in Beliefs. Journal of Mathematical Psychology, 46(1), 88-109.

Hargreaves Heap, S., Rojo Arjona, D., & Sugden, R. (2014). How portable is level-0 behavior? A test of level-k theory in games with non-neutral frames. Econometrica, 82(3), 1133-1151.

Heifetz, A., & Kets, W. (2013). Robust multiplicity with a grain of naïveté. (No. 1573). Discussion Paper, Center for Mathematical Studies in Economics and Management Science.

Ho, T.-H., Camerer, C., & Weigelt, K. (1998). Iterated dominance and iterated best-response in experimental p-beauty contests. American Economic Review, 88 (4), 947-969.

Hommes, C., & Zhu, M. (2014). Behavioral learning equilibria. Journal of Economic Theory, 150, 778-814.

Hommes, C., Sonnemans, J., Tuinstra, J., & Van de Velden, H. (2005). Coordination of expectations in asset pricing experiments. Review of Financial Studies, 18(3), 955-980.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. Econometrica, 263-291.

Kets, W. (2012). Bounded reasoning and higher-order uncertainty, mimeo: http://ssrn.com/abstract=2116626

Kets, W. and A. Sandroni (2015). Challenging Conformity: A Case for Diversity, mimeo.

Keynes, J. M. (1936), The General Theory of Interest, Employment and Money, London: Macmillan.

Kneeland, T. (2015). Identifying higher-order rationality. Econometrica, 83(5), 2065-2079.

Knoepfle, D. T., Wang, J. T. Y., & Camerer, C. F. (2009). Studying learning in games using eye-tracking. Journal of the European Economic Association, 7(2-3), 388-398.

Kocher, M. G., & Sutter, M. (2005). The decision maker matters: Individual versus group behaviour in experimental beauty-contest games. Economic Journal, 115(500), 200-223.

Lambsdorff, J. G., Schubert, M., & Giamattei, M. (2013). On the role of heuristics—Experimental evidence on inflation dynamics. Journal of Economic Dynamics and Control, 37(6), 1213-1229.

Ledoux, Alain (1981), Concours résultats complets. Les victimes se sont plu à jouer le 14 d'atout, Jeux & Stratégie 2, No. 10, 10-11.

Levine, D. K. (2012). Is Behavioral Economics Doomed?: The Ordinary Versus the Extraordinary. Open Book Publishers.

Levitt, S. D., & List, J. A. (2007), What do laboratory experiments measuring social preferences reveal about the real world?. Journal of Economic Perspectives, 21(2), 153-174.

Liberman, V., Samuels, S. M., & Ross, L. (2004). The name of the game: Predictive power of reputations versus situational labels in determining prisoner's dilemma game moves. Personality and Social Psychology Bulletin, 30(9), 1175-1185.

Lucas, R. E. (1972). Expectations and the neutrality of money. Journal of Economic Theory, 4(2), 103-124.

Mauersberger, F. (2016). Monetary Policy Rules in a Non-rational World: A Macroeconomic Experiment. Columbia Working Paper Series.

McKelvey, R. D., & Palfrey, T. R. (1995). Quantal response equilibria for normal form games. Games and Economic Behavior, 10(1), 6-38.

Milgrom, P., & Roberts, J. (1991). Adaptive and sophisticated learning in normal form games. Games and Economic Behavior, 3(1), 82-100.

Mitzkewitz, M., & Nagel, R. (1993). Experimental results on ultimatum games with incomplete information. International Journal of Game Theory, 22(2), 171-198.

Morris, S., & Shin, H. S. (1998). Unique equilibrium in a model of self-fulfilling currency attacks. American Economic Review, 88(3), 587-597.

Morris, S., & Shin, H. S. (2002). Social value of public information. American Economic Review, 92(5), 1521-1534.

Moulin, H. (1979). Dominance solvable voting schemes. Econometrica, 47(6), 1337-1351.

Moulin, Hervé (1986). Game Theory for Social Sciences, New York University Press.

Müller, J., & Schwieren, C. (2011). More than meets the eye: an eye-tracking experiment on the beauty contest game. mimeo.

Nagel, R. (1993). Experimental results on interactive competitive guessing, Discussion Paper No. B-236, University of Bonn, April 1993.

Nagel, R. (1995). Unraveling in guessing games: An experimental study, American Economic Review, 85(5), 1313-1326.

Nagel, R. (1998). A Survey on Experimental "Beauty-Contest Games": Bounded Rationality and Learning, in Games and Human Behavior, Essays in Honor of Amnon Rapoport. Eds. D. Budescu, I. Erev, and R.Zwick. Publisher: Lawrence Erlbaum Associates, Inc., New Jersey, pp.105-142

Nagel, R., Brovelli, A., Heinemann, F., & Coricelli, G. (2015). Neural correlates of risk and strategic uncertainty, mimeo.

Nagel, R., Bühren, C., & Frank, B. (2016). Inspired and Inspiring: Hervé Moulin and the discovery of the beauty contest game, forthcoming in: Mathematical Social Sciences.

Nagel, R., & Selten, R. (1997). 1000 DM zu gewinnen. Spektrum der Wissenschaft (German issue of Scientific American), p. 10.

Niederle, M. (2015). Intelligent Design the Relationship of Economic Theory to Experiments: Treatment driven Experiments. In Handbook of Experimental Economic Methodology edited by Guillaume R. Fréchette and Andrew Schotter.

Nimark, K.P. (2011). Dynamic Higher Order Expectations, working paper.

Nobuyuki, H., Sutan, A., & Willinger, M. (2016). The strategic environment effect in beauty contest games, mimeo.

Paclet, Philippe (1983). La page du matheux (ludique). Questions subsididiaires, Jeux & Stratégie 4, No. 23, 86-87.

Pearce, D. G. (1984). Rationalizable strategic behavior and the problem of perfection. Econometrica, 52(4), 1029-1050.

Pfajfar, D., & Zakelj, B. (2011). Inflation expectations and monetary policy design: Evidence from the laboratory. http://ssrn.com/abstract=1332982

Rapoport, A., Guyer, M., & Gordon, D. G. (1976). The 2 x 2 game. University of Michigan Press.

Robinson, S. (2004). How real people think in strategic games. SIAM News, 37(1), 1-3.

Roth, A. (2009). Nagel's guessing/beauty contest game: a famous experiment in game theory. September 24, 2009. http://marketdesigner.blogspot.de/2009/09/nagels-guessingbeauty-contest-game.html

Roth, A. E., Marilda, A., & Sotomayor, O. (1990). Two-Sided Matching. A Study in Game-Theoretic Modeling and Analysis, Cambridge University Press.

Rubinstein, A., (2007). Instinctive and cognitive reasoning: a study of response times. Economic Journal, 117(523), 1243–1259.

Schelling, T. (1960). The Strategy of Conflict. Cambridge, MA: Harvard Univ. Press.

Selten, R. (1991). Evolution, learning, and economic behavior. Games and Economic Behavior, 3(1), 3-24.

Sbriglia, P. (2008). Revealing the depth of reasoning in p-beauty contest games. Experimental Economics, 11(2), 107-121.

Shapiro, D., Shi, X., & Zillante, A. (2014). Level k reasoning in a generalized beauty contest. Games and Economic Behavior, 86, 308-329.

Simon, H.A, and R. D. Bartel (1986). The Failure of Armchair Economics. Challenge Vol. 29, No. 5, 18-25

Simon, H. A. (1996). The sciences of the artificial. MIT Press.

Simonsen, M. H. (1988). “Rational Expectations, Game Theory, and Inflationary Inertia,” in P.W. Anderson, K. J, Arrow, and D. Pines, eds., The economy as an evolving complex system., Vol. 5. Redwood City, DA: Addison-Weslet, 205-241.

Slonim, R. L. (2005). Competing against experienced and inexperienced players. Experimental Economics, 8(1), 55-75.

Stahl, D. O. (1993). Evolution of smart n players. Games and Economic Behavior, 5(4), 604-617.

Stahl, D. O. (1996). Boundedly rational rule learning in a guessing game. Games and Economic Behavior, 16(2), 303-330.

Stahl, D. O. (1998). Is step-j thinking an arbitrary modelling restriction or a fact of human nature?. Journal of Economic Behavior and Organization, 37(1), 33-51.

Stahl, D. O., & Wilson, P. W. (1994). Experimental evidence on players' models of other players. Journal of Economic Behavior and Organization, 25(3), 309-327.

Stahl, D. O., & Wilson, P. W. (1995). On players′ models of other players: Theory and experimental evidence. Games and Economic Behavior, 10(1), 218-254.

Strzalecki, T. (2014). Depth of reasoning and higher order beliefs. Journal of Economic Behavior and Organization, 108, 108-122.

Tang, F. F. (2001). Anticipatory learning in two-person games: some experimental results. Journal of Economic Behavior and Organization, 44(2), 221-232.

Thaler, R. H. (1997). Giving Markets a Human Dimension. Financial Times, June 16. Sec. 6, pp 2-5.

Thaler, R. H. (2015). Misbehaving: The Making of Behavioral Economics. WW Norton & Company.

Tirole, J. (1982). On the possibility of speculation under rational expectations. Econometrica, 50(5), 1163-1181.

Van Huyck, J. B., Battalio, R. C., & Beil, R. O. (1991). Strategic uncertainty, equilibrium selection, and coordination failure in average opinion games. Quarterly Journal of Economics, 106, 885-910.

Varian, H. R. (2002). Economic Scene; You've seen the movie. Now just exactly what was it that John Nash had on his beautiful mind? NY times, April 11, 2002.

Weber, R. A. (2003). ‘Learning’ with no feedback in a competitive guessing game. Games and Economic Behavior, 44(1), 134-144.

Woodford, M. (2003). Interest and prices: Foundations of a theory of monetary policy. Princeton university press.

Woodford, M. (2013). Macroeconomic analysis without the rational expectations hypothesis. Annual Review of Economics, 5(1), 303-346.

Appendix

I. Moulin (1986, p. 71ff): Sophisticated and Perfect Equilibrium

1. SUCCESSIVE ELIMINATION OF DOMINATED STRATEGIES

A dominating strategy is the most advantageous non-cooperative strategy one can employ no matter how much that player knows of the other players' utilities. By contrast, the scenarios proposed in this chapter rely on the complete information assumption. That is, each player is aware of the entire normal form of the game, including other player's preferences. Adhering to this assumption, noncooperating players mutually anticipate each other's strategic choices. Each player expects all other players to eliminate their dominated strategies. This, in turn, may generate new dominated strategies, and so on.

Example 1. Guess the average.

Each player picks an integer between 1 and 999. Let

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$$

be the average answer. The winners are those players whose ballot is closest to 2/3 $\bar{x}$.

Every strategy above 666 is dominated (by 666). Thus if all players use undominated strategies only, each strategy set shrinks from {1, ..., 999} to {1, ..., 666}. The argument can be repeated. Every strategy above 444 is now dominated (by 444), so nobody will use any such strategy, and so on. Finitely many repetitions of the argument force the conclusion that each player chooses $x_i = 1$.

Definition 1 (*Moulin [1979]*).

In the n-player game $G = (X_i, u_i, i = 1, \ldots, n)$, *the successive elimination of dominated strategies is made up of the sequences*

$$X_i = X_i^0 \supset X_i^1 \supset \ldots \supset X_i^t \supset X_i^{t+1} \ldots \text{ all } i = 1, \ldots$$

where $X_i^{t+1} = D_i(u_i; X^t)$.

We denote $X_i^\infty = \bigcap_{t=0}^{\infty} X_i^t$. *We shall say that G is dominance solvable if* X_N^∞ *is nonempty and Player i's utility* u_i *does not depend on* x_i *on* X_N^∞:

$$u_i(x_i, x_{-i}) = u_i(y_i, x_{-i}) \text{ all } x_i, y_i \in X_i^\infty, \; x_{-i} \in X_{-i}^\infty$$

In this case we call X_N^∞ the set of ***sophisticated equilibria*** *of G.*

Typically if X_i^∞ is a singleton for all i = 1, ..., n, then G is dominance solvable. The behavior of sophisticated players is both static and decentralized. The game is played once after each player has performed, independently, the successive elimination of dominated strategies; when this algorithm converges (i.e., the game is dominance solvable) our players eventually pick an unambiguous equilibrium strategy. In game (X_i, u_i, i = 1, ..., n) all Player i's strategies are equivalent to him. True, this does not imply that they are equivalent to other players (see Example 3, Chapter 3) yet, in general (e.g., when utility functions are one to one on X_N), we can expect the sophisticated equilibrium to be a singleton. In dominance solvable games, sophisticated behavior is essentially deterministic.

Our next example displays implicit cooperation at sophisticated equilibrium.

Example 2. Plurality voting with ties (Farqharson [1969]).

Among three candidates {a, b, c}, a society 1, 2, 3 must elect one. The voting rule is plurality voting and Player 1 breaks ties. In other words, the strategy sets (or message set) are $X_1 = X_2 = X_3 = \{a, b, c\}$, and, if the agents cast the votes (x_1, x_2, x_3), the elected candidate is

$$\pi(x_1, x_2, x_3) = x_2 \qquad \text{if } x_2 = x_3$$

$$= x_1 \qquad \text{if } x_2 \neq x_3$$

Suppose now that the utility of players for the various candidates display a Condorcet effect,

$$u_1(c) < u_1(b) < u_1(a)$$

$$u_2(b) < u_2(a) < u_2(c)$$

$$u_3(a) < u_3(c) < u_3(b)$$

Player 1 has a dominating strategy (vote for a) hence his ballot is predictable at once. Player 2 and 3, however, have only one dominated strategy, namely, to vote for their least-preferred candidate. Thus, after one round of elimination, we have

$$X^1_1 = \{a\},\ X^1_2 = \{a, c\},\ X^1_3 = \{b, c\}$$

In game $(X^1_i, u_i\pi, i = 1, 2, 3)$ we note that outcome b cannot be elected any longer, hence Player 3's strategy b is <u>now</u> dominated by his strategy c. By the same token, Player 2's strategy a is now dominated by his strategy c. Thus, after two rounds of elimination, outcome (a, c, c) emerges as the sophisticated equilibrium. Player 1's privilege as the tie breaker yields eventually his worst outcome!

How realistic is the strategic behavior implied by the concept of sophisticated equilibrium? It presupposes that each and every Player either computes <u>all</u> dominated strategies of all players or performs the successive elimination as long as necessary, on the assumption that every other player is equally patient. The first point does not cause much trouble. In fact, when the game is dominance solvable, nothing would be changed if the elimination process was to take place successively, in arbitrary order; the players could also drop only part of the dominated strategies. These robustness properties, however, are not easy to prove (see Lemma 1 below).

The second point raises a more serious objection. As Example 1 above makes clear, the successive elimination can be arbitrarily long. In practice, most players do not perform the elimination forever out of fear that other players are not rational enough to do so. This is what experimental evidence of the "guess the average" game suggests. That is, if 20 players are involved, you do not expect that all of them will perceive the geometrical shrinking of strategy sets. In fact, they do not, and the winning guess usually lies between 100 and 200.

Our last example shows that elimination of dominated strategies can yield a deterministic noncooperative outcome even when the game is not dominance solvable.

II. Jeux & Stratégie n°10 - août-septembre 1981:

Il s'agissait de donner un nombre entier positif inférieur à un milliard. Nous ferions la moyenne des nombres proposés par les ex aequo. Les gagnants étant, dans l'ordre, ceux qui seraient les plus proches des 2/3 de cette moyenne.

Il fallait bien, sûr avoir de la chance pour décrocher le premier prix, mais nous pensons que tous ceux qui ont fait un bon raisonnement doivent se trouver parmi les 350 gagnants. Etant donné la question, le bon sens interdisait en effet de proposer un nombre important puisque la première réflexion incitait à choisir un nombre le plus petit possible. Mais, à l'opposé, était-il raisonnable de choisir « 1 » comme l'ont fait certains candidats ? (sauf si on espérait n'être opposé qu'à un seul autre ex aequo). Finalement, le résultat nous a paru « logique ».

La moyenne s'est établie à 134 822 738,26 dont il fallait donc jouer les 2/3 soit 89 881 825,51. Le candidat le plus proche nous a proposé 89 793 238. Le plus petit nombre récompensé a été 10 457 (350e), le plus grand 179 451 328 (312e).

Voici enfin la liste des 350 gagnants. En effet, devant le nombre et la qualité des réponses nous avons décidé d'ajouter 250 nouveaux prix. En plus de la liste publiée dans le n° 8, seront ainsi distribués quatre excellents jeux édités par Interlude : Skirrid, Zakhia, Quad-ominos et Brain Trainer. (Nous avons dû recourir au tirage au sort pour départager quelques ex-aequo. Heureusement, ce cas ne s'est pas produit avant la 32e place !)

Jeux & Stratégie n°20 - avril-mai 1983:

Question subsidiaire :
Incrivez ici un nombre entier positif inférieur à un milliard.
Cette réponse servira à la fois au classement intermédiaire (prix spécial) et au classement final, mais selon deux modalités différentes expliquées page 13.

Jeux & Stratégie n°22 - août-septembre 1983:

Etait-ce plus facile que l'année dernière ? Ou bien avez-vous été plus courageux... et plus perspicaces ? Toujours est-il que cette fois, ce sont près de 10 000 réponses que nous avons reçues. Et sur les 9 298 arrivées dans les délais, exactement 6539 ont passé correctement les premières épreuves, participant ainsi au classement du **Prix Spécial.**

Et, en ayant joué 33 289 564 à la question subsidiaire, c'est Claude Tanagro, de Sète qui remporte la console Colécovision (CBS).

Après toutes les questions, il restait encore 3 502 bulletins justes et parmi ceux-ci, seulement 604 qui n'obtenaient pas le « top » aux épreuves de départage, qui ont encore mal résisté à vos recherches. Comment donc faire pour vous coller ?

Finalement, 2 898 bulletins ont donc dû être classés par la question subsidiaire. Les deux tiers de la moyenne des nombres proposés s'établissaient à 67 329 453. La réponse la plus proche était celle d'Armand GUT, de Toulon, qui, avec 67 373 773, gagne le voyage au Mexique.

Voici à présent la liste complète des trois cents gagnants du concours J & S 83.

III. Jeux & Stratégie n°22 - août-septembre 1983:

la page du matheux (ludique)

questions subsidiaires

L'épreuve que nous avions choisie à deux reprises pour départager les nombreux ex aequo de nos concours, nous a valu un abondant courrier. Sans prétendre fournir LA solution, puisque la question était justement conçue pour ne pas en admettre, voici nos réflexions sur le sujet.

Rappelons cette question, pour ceux qui n'ont participé ni au concours 81, ni à l'édition 83 où, tirant sa nouveauté de sa répétition, cette même question fut posée de nouveau.

« Les concurrents devront citer un nombre entier entre un et un milliard. Le gagnant sera celui dont le nombre se rapproche le plus des 2/3 de la moyenne des nombres proposés par tous les ex aequo ».

Si vous n'avez jamais réfléchi à cette question, nous vous engageons à le faire, et, pourquoi pas, à proposer une réponse avant de continuer. Le sentiment de vertige qu'elle procure aux premiers contacts est loin d'être désagréable.

C'était une bonne question. D'abord, ceux des concurrents pouvant disposer d'un ordinateur n'étaient pas trop avantagés, l'analyse qu'on pouvait en faire ne demandant a priori que des calculs simples de moyenne pour lesquels une calculette était suffisante. En revanche, elle accordait de meilleures chances de gain à celui qui saurait voir plus loin que ses adversaires. Pour prendre un exemple extrême, supposons que le groupe des ex aequo soit composé de 2 999 singes et d'un être humain : en jouant 333 333 333, ce dernier rend maximales ses possibilités de gagner, car si les singes jouent au hasard, la moyenne sera probablement proche d'un demi-milliard.

En dernier lieu, cette question n'était évidemment pas « déterministe » : car si on peut espérer délimiter des zones préférentielles où la « bonne réponse » a quelque probabilité de se trouver, ce choix n'est malgré tout ni unique, ni sûr, ni précis.

D'ailleurs, est-il vraiment possible de délimiter par l'analyse ces fameuses zones préférentielles ? J'en ai personnellement longtemps douté, mettant en avant la situation paradoxale que l'existence de telles zones viendrait à créer. Si elles existaient, elles seraient, en théorie, accessibles à tous et perdraient par là même leur prépondérance. Mais c'était ne vouloir penser qu'à l'aspect purement mathématique du problème, ignorer qu'en l'occurence, le plus important est ce qui se passe dans la tête des gens, là où il est bien illusoire de s'attendre à l'unanimité et à l'uniformité.

L'analyse qui va suivre repose sur l'axiome qu'il est possible de classer, raisonnablement, les comportements en quelques groupes homogènes simples qui, par leur antagonisme, donneront naissance à un certain équilibre. Elle ne prétend pas détenir la vérité absolue, dont la remarque ci-dessus nie par ailleurs l'existence. Mais elle a l'avantage de s'appliquer avec un assez bon succès aux deux cas de figures qu'il nous a été donné d'observer. Précisons enfin qu'elle a été menée après coup, en examinant les résultats du concours 83 (voir encadré).

Elle part de la remarque qu'aucun des candidats n'aura pu manquer de faire : pour espérer gagner il ne faut en aucun cas jouer un nombre supérieur à 666 M (1). Or, si tous les nombres cités sont inférieurs à 666 M, les 2/3 de leur moyenne feront au (grand) maximum 444 M. Cette remarque est identique à la première, et là encore, cette idée n'aura pu échapper à personne. Une fois que le pli a été pris, rien n'est plus tentant que de recommencer mécaniquement le même raisonnement ce qui porte inéluctablement à 1.

Faut-il donc choisir 1 ? Certains le prétendront, avec insistance parfois, s'il faut en croire certaines lettres ou coups de téléphone reçus à la revue.

Mais, pratiquement, pour que la moyenne s'établisse à 1, ou même à moins de 100 000, il faut absolument qu'aucun concurrent n'ait joué « grand ». Il suffit d'un distrait, d'un fou, ou d'un singe, qui ait proposé ne serait-ce que 333 M pour que la moyenne monte au-dessus de 100 000 (2). Et puis, il y a les petits malins, ceux qui jouent en groupe, qui n'auront pas mis longtemps à réaliser qu'il était possible de faire monter « artificiellement » la moyenne en sacrifiant un ou plusieurs de leurs membres sur l'autel des grands nombres. Une dizaine de propositions à 999 999 999 suffisent déjà à assurer une moyenne supérieure à 3 M. Les résultats font apparaître une centaine de propositions « à fonds perdu », au-dessus de 666 M et en particulier six 999 999 999.

Est-ce à dire qu'il y a eu tricherie ou même, comme l'imagine un de nos lecteurs, manipulation sur grande échelle des résultats ?

Je ne le pense pas. L'hypothèse la plus vraisemblable (3) est que de nombreux petits groupes ont eu la même idée.

Le joueur isolé était-il fondamentalement désavantagé ? Encore une fois non, puisqu'une analyse réaliste de la situation (vos adversaires ne sont pas des singes) montre qu'il est tout à fait improbable que de telles manipulations ne se produisent pas. Sur quelles échelles ont-elles lieu ? les joueurs groupés le savent-ils mieux qu'un joueur isolé ? A la limite, on pourrait même soutenir qu'un concurrent isolé, jouant bien, s'en trouve avantagé, puisque pouvant en profiter sans avoir à partager à la fin.

On voit donc déjà se délimiter deux groupes de concurrents. Le premier, A, représentant un pourcentage p dés 3 000 ex aequo, est le groupe des « irréductibles des petits nombres ». Nous fixerons (arbitrairement) la moyenne de ces petits nombres à 2,5 M. Le second groupe, B (q % de la population), est celui des « sacrifiés » qui vont jouer gros, et même très gros (quittes à se sacrifier...): moyenne aux alentours de 900 M.

Le reste de la troupe sait qu'il ne faut pas jouer si petit que ça ; mais jusqu'où faut-il remonter ? Pas au-dessus de 666 M en tous cas. Dans une certaine mesure, on pourrait même dire qu'il est de plus en plus difficile de remonter. Ce qui amène à formuler une hypothèse a priori : le gros de cette troupe se répartira entre 5 M et 666 M suivant une loi de probabilité ayant pour densité une droite décroissante, cette densité étant nulle en 666 (fig. 1). Hypothèse simplificatrice sans nul doute, mais qui ne sonne pas si mal. La moyenne de ce groupe sera alors

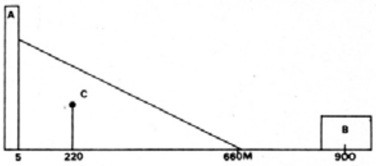


Figure 1 : *une tentative pour isoler parmi les concurrents 3 groupes homogènes. C, réparti sur un triangle, a sa moyenne au niveau du centre de gravité.*

(1) M = millions.

(2) Fixons à 3 000 le nombre des ex aequo.

(3) Hypothèse confortée par l'examen de l'origine géographique de ces propositior « malhonnêtes ».

proche de 220 M (c'est le théorème du centre de gravité d'un triangle). Nous appelerons C ce groupe, qu'il serait absurde d'assimiler à l'ensemble des concurrents n'appartenant ni à A, ni à B, puisqu'il y aura bien un groupe D de joueurs juchés sur le même perchoir que nous (sans pour cela avoir suivi obligatoirement le même raisonnement) et dont nous appelerons r le pourcentage [ce qui laisse $1 - (p + q + r)$ au groupe C].

On peut calculer la moyenne des groupes A, B et C.

$m \simeq (220 - 217p + 700q - 220r) \times 1/(1 - r)$

Introduisons maintenant le groupe D en deux étapes. Une partie, d'importance $r - a$, jouera « bêtement » les 2/3 de m, sans se rendre compte qu'ainsi la moyenne sera rabaissée à $m' = m(1 - r) + 0{,}66.M(r - a) \times 1/(1 - a)$

La deuxième partie, de pourcentage a, jouera au contraire le nombre x, solution de l'équation : $x = 2/3(m'(1 - a) + xa)$

Pour tenir compte des variations vers le bas que causent le fait de jouer plus petit que la moyenne escomptée.

Calculs faits, on arrive à la formule

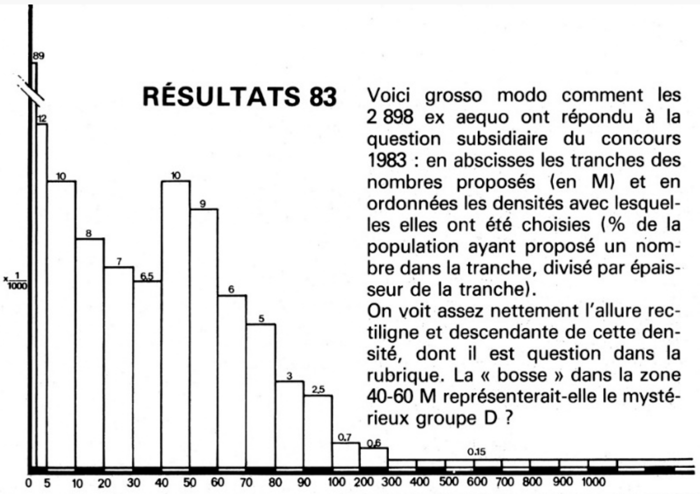
$x = (220 - 217p + 700q - 220r) \times (1 - 0{,}33r - 0{,}66a) \times 1/(1{,}5 - a)$

Belle formule direz-vous, mais dans laquelle entre une bonne part d'arbitraire. C'est certain. Mais ses qualités sont d'une part d'exister, d'autre part d'être capable de délimiter une zone (disons de 50 à 150 M) assez stable vis-à-vis des variations envisageables dans les paramètres introduits. (Il faut quand même limiter ces variations à une fourchette raisonnable : B ne sera pas très grand, tout comme D).

Le tableau de la fig. 2 montre les résultats des calculs pour différentes valeurs des paramètres p, q, r et a. Il montre aussi, en dernière colonne, ceux qu'on obtiendrait dans le cas où la moyenne du groupe C serait placée à 180, soit un peu moins du tiers de 660.

p	q	r	a	x	x'
10	1	10	5	117	97
10	3	5	1	136	114
30	1	10	1	89	74
30	4	5	5	112	95
20	0,5	15	10	92	76
40	10	2	0	130	115
30	1	5	0	98	81
5	2	5	3	138	115
10	0,5	20	10	97	80
10	1	1	0	134	111

Figure 2 : *où faut-il jouer ? Différents résultats suivant les différentes valeurs attribuées aux paramètres p, q, r et a (exprimés en %). Les deux colonnes de x expriment des millions. La première correspond au cas où la moyenne du groupe C est estimée à 220 M, la seconde, à 180 M.*



RÉSULTATS 83

Voici grosso modo comment les 2 898 ex aequo ont répondu à la question subsidiaire du concours 1983 : en abscisses les tranches des nombres proposés (en M) et en ordonnées les densités avec lesquelles elles ont été choisies (% de la population ayant proposé un nombre dans la tranche, divisé par épaisseur de la tranche).

On voit assez nettement l'allure rectiligne et descendante de cette densité, dont il est question dans la rubrique. La « bosse » dans la zone 40-60 M représenterait-elle le mystérieux groupe D ?

Entre 50 et 150 M, pourquoi ne pas choisir le milieu : 100 M ? Rappelons qu'en 1981, c'est en jouant 90 M qu'on gagnait, puisque la moyenne globale des nombres choisis fut à peu près 134,8 M. La formule que nous proposons, qui tire sa raison d'être d'un péché de présomption, à savoir l'affirmation que seul un petit nombre de concurrents y aurait accès, résiste donc bien à l'expérience de 1981.

Qu'en est-il pour l'édition 83 ? On pouvait bien penser qu'un grand nombre de concurrents seraient au courant du résultat 81 et que les comportements en seraient modifiés.

Essayons d'adapter le raisonnement mené plus haut à cette nouvelle situation. A nouveau 4 groupes se profilent : le groupe A (pourcentage p) de ceux qui jouent comme si de rien n'était (ils ne connaissent pas le résultat de 81, ou ils jugent impossible d'en tenir compte) ; le groupe B des « sacrifiés » (pourcentage q) ; le groupe C de ceux qui vont jouer entre 5 et 135 M ; et le groupe D, restreint comme il se doit, de ceux qui voient plus loin que les autres. Nous ne faisons pas apparaître de groupe particulier pour ceux qui vont jouer petit, ou plutôt considérons-nous que ceux-là sont à englober dans le groupe A, étant donné que, s'il est une leçon à tirer de l'expérience 1981, c'est bien qu'on ne peut espérer gagner en jouant moins de 5 M.

Que peut-on dire du groupe C ? Dans un premier temps on peut le traiter comme précédemment : il sera distribué selon une loi dont la densité est une droite décroissante sur l'intervalle de 5 M à 135 M. Ce qui donne pour ce groupe une moyenne de 45 M. Mais en deuxième analyse, on a le droit de penser que cette moyenne pourrait être réestimée à la hausse étant donné que ce groupe est constitué de participants qui auront, au vu des résultats précédents, moins tendance à jouer bas.

A partir de quoi, on peut écrire une formule analogue à celle de la 1re partie. Différents choix des paramètres donnent des résultats s'étalant de 40 à 60 M, quand la moyenne du groupe C est estimée à 45 M, et de 55 M à 80 M quand on la relève à 60 M.

Le vainqueur de cette année, Armand Gut, qui aime bien les 3 et les 7, avait proposé 67 373 773. A l'intérieur d'une fourchette 40-80 M, les résultats du tableau cernent encore une fois la réalité. Encore que cela soit moins spectaculaire que la 1re fois.

Répétons-le, il ne s'agit que d'une façon d'aborder le problème parmi d'autres, trop fortement influencée sans doute par des observations faites a posteriori.

L'hypothèse émise sur la densité du gros de la troupe est-elle naturelle ? Ou ne nous est-elle venue en tête que parce que nous avions les résultats sous les yeux ? J'avoue ne plus tellement m'y retrouver.

Il serait très intéressant de pouvoir disposer d'un grand nombre de témoignages sur cette question. Aussi, nous vous lançons un appel : écrivez-nous, expliquez-nous ce qui vous est passé par la tête au moment de répondre. Ce sera peut-être l'occasion de reprendre le sujet de manière tout à fait différente. Faites-nous part de vos réflexions les plus secrètes. De toute façon, n'ayez crainte, la question est définitivement brûlée : plus jamais nous ne pourrons l'utiliser dans un concours de *J & S*. Pourquoi d'ailleurs ne pas proposer vous-mêmes la question subsidiaire du prochain concours. Toutes les suggestions seront les bienvenues !

Philippe Paclet ●

IV. Back cover of Journal of Political Economy Vol. 100, No.5, October 1992:

Games with Asymmetrical Intelligence

"I knew [a school-boy] about eight years of age, whose success at guessing in the game of 'even and odd' attracted universal admiration. This game is simple, and is played with marbles. One player holds in his hand a number of these toys, and demands of another whether that number is even or odd. If the guess is right, the guesser wins one; if wrong, he loses one. The boy to whom I allude won all the marbles of the school. Of course he had some principle of guessing; and this lay in mere observation and admeasurement of the astuteness of his opponents. For example, an arrant simpleton is his opponent, and, holding up his closed hand, asks, 'Are they even or odd?' Our school-boy replies, 'Odd,' and loses; but upon the second trial he wins, for he then says to himself: 'The simpleton had them even upon the first trial, and his amount of cunning is just sufficient to make him have them odd upon the second; I will therefore guess odd';—he guesses odd, and wins. Now, with a simpleton a degree above the first, he would have reasoned thus: 'This fellow finds that in the first instance I guessed odd, and, in the second, he will propose to himself, upon the first impulse, a simple variation from even to odd, as did the first simpleton; but then a second thought will suggest that this is too simple a variation, and finally he will decide upon putting it even as before. I will therefore guess even';—he guesses even, and wins. Now this mode of reasoning in the school-boy, whom his fellows termed 'lucky,'—what, in its last analysis, is it?"

"It is merely," I said, "an identification of the reasoner's intellect with that of his opponent."

"It is," said Dupin; "and, upon inquiring of the boy by what means he effected the *thorough* identification in which his success consisted, I received answer as follows: 'When I wish to find out how wise, or how stupid, or how good, or how wicked is any one, or what are his thoughts at the moment, I fashion the expression of my face, as accurately as possible, in accordance with the expression of his, and then wait to see what thoughts or sentiments arise in my mind or heart, as if to match or correspond with the expression.' This response of the school-boy lies at the bottom of all the spurious profundity which has been attributed to Rochefoucault, to La Bougive, to Machiavelli, and to Campanella."

[Edgar Allan Poe, "The Purloined Letter," in *The Complete Tales and Poems of Edgar Allan Poe* (New York: Random House, 1938), 215–16]

(Suggested by Martin Hellwig)

Appendix V: Personal comments

Alain Ledoux´s e-mails (translated from French to English):

Subject: Psycho-statistique

Date: Thu, 9 Jul 2009 11:32:14 +0200

To Pr Dr Björn Frank

Dear Sir Frank,

[…] I was very amused to answer your questions proposed on the site of ChessBase. Because I well think to be ... the inventor [of the game]!

I had this idea in 1981 to determine the winners in the first competition organized by the magazine *Jeux & Stratégie* I created a year before and for which I was the editor. I am sincerely flattered that this issue has been interesting to a university twenty-eight years later! Were you at the time a reader of this magazine, and were you thereby inspired or did this idea come to you independently? Today I am almost retired and *Jeux & Stratégie* having long since disappeared, I am always curious about everything regarding the ten years of the great adventure of this magazine which remains my pride.

For the record, the result was 8, 979 (a comma, as we had requested an integer between 1 and 1000000000) ... Two years later, the same question were resting our readers and response was then 0, 06 737 ...

I look forward to the results and I would be really glad you tell me a little about your interest in this question which at the time I had pompously described as "psycho-statistic".

Best regards,

Alain Ledoux

References :

First results: J & S # 10 (August, September 1981), the question was asked individually by post to tie the contest proposed in No. 8.

Second result: J & S # 22 (August, September 1983).

10 March 2010:

Dear Rosemarie Nagel,

[…]

You should receive the famous No. 10 *Jeux & Stratégie*. You will discover on page 11 (very discrete but that seems to be the first) description of the game "Guess 2/3 of the average". I honestly cannot tell you how this idea came to me, just that I was really anxious to find a way to separate the many ties to our contest by means that do not only appeal to chance but that requires some thought. But which frankly surprises me is that I chose the 2/3 coefficient quite intuitively (and why not ¾ for example?) And even today, it is this value that remains employed. Surprising indeed, the fact that I never heard about this for twenty-eight years, until it was proposed in June on the site of ChessBase (even in 1981, I have not have heard any comment about that subject...).

How I came to the game? My father taught me when very little to play checkers and chess. I then became a passionate chess player, and I still play in competition, even if it is with the results of a very average amateur ... After studying engineering, I started a career as a journalist entering the magazine Science & Vie (1973). And I naturally created a section of chess. As it worked well, I then launched a section of go, a game that was beginning to be known in France (of course, this time it was not me who was writing). New success. And so in a few years, the "games" of this scientific magazine has grown to reach a dozen pages (math games, critique new games ...). The late 1970s was extraordinarily innovative in the entertaining area: first war games, the first games of role, first machines to play chess, birth of personal computing and thus computer games ... so I suggested to the *Science & Vie* editor to launch an exclusively devoted to games magazine. This was the birth of *Jeux & Stratégie* (January 1980). Looking back, it was really innovative. There was nothing comparable in the world (except maybe English *Games & Puzzles*, excellent but still very different). And for ten years (until July 1989 when the publisher decided to stop the publication that was losing money because of the lack of advertising revenue despite a still respectable spread over 45 000 copies), it was an amazing adventure, between a small team of as creative collaborators (2) fans and readers of an incredible proximity.

I've already said, the copy you will receive is really not the most successful. I'll get if I can send you another (for example, with the results of the second question "Guess 2/3 of the average"). I must also check one small detail. In his note on this story Professor Björn Frank reports that I had called such questions "psycho-statistics", but reports that it seems that I did not write this in the journal. But I'm absolutely sure about it. But I do remember when ...

Last point. I have a hypothesis to explain why it was said that this issue had appeared in the French version of Scientific American. As I have said *Jeux & Stratégie* was published by the Science & Vie editor, a magazine which can be considered (not necessarily rightly) as the "equivalent" French *Scientific American*, and of course not his French "version", but perhaps this is the origin of confusion.

That's the story …

Alain Ledoux

From: Jean François Laslier
Date: 2016-05-12 17:59 GMT+02:00
Subject: a few lines

Dear Rosemarie

I was a student at the ENSAE (Ecole Nationale de la Statistique et de l'Administration Economique) a French « Grande Ecole » in 1981-1984. Hervé Moulin was giving the course on Game Theory. I very well recall the day he talked about the "two-thirds of the mean" game. He mentioned that the game was found in "Jeux et Stratégies" as a tie-breaking question, and that it was proposed two years in a row. This magazine was popular among students.

So this 2/3 of the mean game stayed in my mind as a proof (among others) that one cannot restrict the study of interactive situations to equilibrium computations. This is why I decided to do my PhD on the micro economic models of bounded rationality and self-organization on which Jacques Lesourne was working with Gilbert Laffond (Hervé's first PHD student).
These topics were of course very difficult to publish in English at that time, and most papers were published in the journal Economie Appliquée. But two books are available in English:
Jacques Lesourne (1991) Economie de l'ordre et du désordre. Paris : Economica. English translation: The Economics of Order and Disorder, Oxford: Clarendon Press, 1992.
Jacques Lesourne, André Orléan and Bernard Walliser (eds.) (2002) Leçons de microéconomie évolutionniste. Paris : Odile Jacob. English translation: Evolutionary Microeconomics. Heidelberg : Springer, 2006.

Jean-Francois Laslier
Paris School of Economics

From: Martin Hellwig
Date: 2016-06-09 12:30 GMT+02:00
Subject: Re: a history on the beauty contest game.. in honour of Moulin's 65's birthday

Dear Rosemarie Nagel,

thank you for your inquiry. I got acquainted with the game through conversations with Roger Guesnerie and decided to try it in a game theory class in Basel. (You might also want to ask Tilman Börgers who was involved in the course as well.) At the time I was much impressed that (i) there several students who found the "correct" solution and (ii) the winner who had given an "incorrect" number (11 if I remember correctly) explained that he also knew the correct solution but decided that many others did not and therefore...

Concerning the quote from Poe, the title was chosen by the JPE, not by me.

Moreover, you might find it worthwhile to read on in the story by Poe. The quote comes from a conversation between the person who poses as telling the story and Dupin. The conversation concerns the question of what the minister, a sinister person who has stolen an incriminating letter, might be thinking and how he might proceed. At some point Dupin refers to the minister as a poet. The teller of the story interjects "I thought he was a mathematician, and the poet was his brother!" To which Dupin replies "There is only one person, and he is both a mathematician and a poet. And that is what makes him so dangerous. If he was only a mathematician, he would not have any real imagination, but being a poet as well, he has..." I am quoting from memory, so the words may differ, but the gist is there. The message of course is one that ought to give food for thought to mathematical economists.

Best regards,
Martin Hellwig
From: Eric van Damme

Date: 2016-05-25 10:31 GMT+02:00
Subject: Re: a history paper on the beauty contest game. a question

Dear Rosemary,

My memory is quite imperfect. I will reflect on the matter a bit more. I think I read a small book by Herve when I was a PhD student. I think it was an English translation from a French one. If I recall well Herve had a concept related to dominance solvability and in that context had an example or an exercise in which the questing game was discussed.

This must have been around 1980, hence, the only experiments then were about market games and the only connection between experiments and game theory was through Reinhard Selten. For sure, Alvin Roth was mainly working on axiomatic bargaining at the time; maybe he was beginning to do experiments on cooperative bargaining theory. Hence, it did not cross my mind to do experiments with this game at the time. As we now know, nobody thought about it at the time. Where there experiments to test non-cooperative concepts in non-zero sum games (non-market games) before Werner's ultimatum game experiments?

Later the rationalizability concepts drew much attention, but even then at first nobody did serious experiments. When your work appeared, I guess that many of us might have regretted not having seen the goldmine in front of us.
Best,
Eric

From: Tore Ellingsen
Date: 2016-06-15 13:08 GMT+02:00
Subject: Re: history on the BC game... in honour of Moulin's 65 birthday

Hi Rosemarie

Thanks, this is interesting! I'm pretty sure I first saw the game in a lecture by Guesnerie at LSE. Are you sure you heard him after my lecture and not before? If your recollection is correct on this point, then I must have attended an earlier lecture by him. (I arrived at LSE in 1988, so it's quite possible. I think he visited several times. But my sense was I heard it in the fall of 1990.)

I remember that you were enthusiastic about running experiments on this game, but I cannot recall my response. Presumably it was lukewarm; I certainly did not spot the goldmine. The next thing I heard about it, I think, was when Eric van Damme praised the experiment and your model in a conversation I had with him at a conference he organized in January of 1992.

It's also intriguing to think about the game's name. At one level (deep), Selten was obviously right. At another level (much more shallow) BC may be appealing because it points out very clearly a pitfall in this situation: Don't focus on what you think is desirable or somehow "right". Focus on what the others are likely to do. For choices of titles, of games and papers, the best strategy in terms of maximizing short-term impact appeals to as many as possible, not necessarily to the greatest connoisseur. However, in the long run, the masses – even within a discipline – are usually wrong. The work of the connoisseurs survives. I will be calling it the Guessing Game more consistently from now on.

All the best
Tore

From: Philippe Paclet
Date: 2016-11-21 11:29 GMT+01:00
Subject: Re: about your article in Jeux & Stratégie 4, No. 23, 86-87.1983 (related to Mr Ledoux)

Dear Prof. Nagel, Dear Rosemary

sure, your mail was a surprise, a very pleasant one indeed, thank you.

It brings me back more than 30 years ago, to the good times I had participating to "Jeux&Strategie" aventure in the early eighties together with a little creative team animated by Alain. I was the "Matheux Ludique".

In 1978, I had passed my doctorate in Pure Mathswith a thesis on Potential Theory in infinite dimensional space (let's say) (Université P.et.M.Curie – Paris) but I did not manage to obtain a stable position in the university or any research laboratory.From 1976 to 1979, I had a grant to do research in PDE and functional analysis at the Scuola Normale Superiore in Pisa with G. Stampacchia and E. de Giorgi. But it was coming to an end and my family situation (a new born baby with a very severe handicap) urged me to abandon my young boy dream of doing research in Maths. So I decided to pass the 'Agregation', a famous 'concours' in France to select teachers for the french lycees, which I won in July 1980.

One obligation which was asked to the laureates of Agregation was to spend one month working 'outside' the education system. A friend of mine, Nicolas Giffard, one our top french chess players in that time, was already in 'J&S" taking care of the chess column and he presented me to Alain who accepted me as a trainee. We got along very well, so I remained more than one month.

My role was to propose curious mathematical problems, variations on known ones or invented. I had also to write these 2 pages section' "le Matheux Ludique" trying to untertain the readers with Mathematics. My model was Martin Gardner, of course. Since my earthly youth, I had always a passion for math and I enjoyed very much to have the opportunity to share this passion. It went on till 1984, if I remember well, because in September 1982, I got a position of Mathematics Professor at the Lycee Chateaubriand, the french lycee in Roma. I tried for a year or 2 to continue with J&S but it was materially difficult (no internet, no email in those times) to send my article, receive the first drafts, do the corrections and so on.

So I concentrated on my teaching responsibilities: I never wanted to teach, but I tried to do it the best I could, tried to transmit my passion to my students.On the whole (I am freshly retired) I think I have been successful on this point, and, in that, my participation to J&S has been very helpful. Not a few of my students have chosen mathematical studies at university, and, the greatest reward for a teacher, have surpassed me.

I do remember very well the "question subsidiaire": when Alain came out with it, out the blues, I immediately found it very, very interesting. For a moment though, I was a bit effrayed that it would not reach the goals we were waiting from that question, i.e. to able to select few competitors among many who had resolved all the preliminary already difficult questions regarding board games, mathematical puzzles, etc.., showing in that reasoning capacities greater than usual. At first glance, it seemed obvious to me that most of them would have chosen 1, leaving us with the initial problem (too many winners), because the first step (to choose a number >666M is a dominated strategy) is rather easy to take,...and once it is taken the infinite geometric descent is wide opened...which is probably true for a mathematical oriented mind, but not for all. I was far from imagine the k-levels... But a few test of the question to random person around convinced me that it might work. And it did, as did the re-proposition of 1983.

I was not at all aware of what important matter this question has become in game theory applied to social science over the years, and the role you had in its development. Thank you very much for sending your historical notes about the subject which I have started to read with some difficulties because I don't grasp firmly all the technical terms of game theory and behavioral science. But you have triggered my curiosity: as I said I am just retired but I still spend many hours with mathematics,re-studying subjects (number theory, probabilities,..) which I have forgotten after so many years of teaching at lycee level, studying things I never had studied before such as statistics (which, in my youth, romanticly, as G.H.Hardy, I looked over with a bit of contempt as all 'Applied Maths'. I have radically changed my mind since).

Your article contains many references; would it be too much to ask you a recommended selection to start with, knowing that the more maths it contains is probably the easier way for me to start the journey?

I am of course very grateful to you for (re)discovering the role J&S played in that story and telling it. As I said before, the original question was definitively Alain's one (and I am pretty sure he didn't have heard about Keyne's beauty contest before) so if you have any doubt about the name it should be given (pbeauty contest is not nice), why not 'Ledoux-Moulin question'?

My role in that story is marginal and I surely did not foresee all the potentiality of the problem. I am glad you read my article with interest and judged it worth of interest, enough at least to re-publish it after Moulin's one. But if it contains, as you say, some insights that have been confirmed later on in a more scientific and structured way, I am curious to fill the 33 years gap and read about the development of the question since Moulin unvealed its academic relevance.

How strange the 2/3 coefficient remained so stable in the experiment. Did anybody study cases, theoretically or experimentally, where p was much nearer to 1, p=0.9? Also, did anybody studied the case n=3 or other small values of n?

This lengthy letter comes to an end. Its length measures the excitement yours provoked in me. No doubt, they will be some more in the future because I will certainly need to ask you some more questions, and, who knows, maybe propose some ideas.

Thank you again.

Philippe Paclet