

Discussion Paper

No. 2019-53 | November 06, 2019 | <http://www.economics-ejournal.org/economics/discussionpapers/2019-53>

Please cite the corresponding Journal Article at

<http://www.economics-ejournal.org/economics/journalarticles/2020-16>

Bounded rationality in Keynesian beauty contests: a lesson for central bankers?

Felix Mauersberger, Rosemarie Nagel, and Christoph Bühren

Abstract

The goal of this paper is to show how adding behavioral components to micro-founded models of macroeconomics may contribute to a better understanding of real world phenomena. The authors introduce the reader to variations of the Keynesian Beauty Contest (Keynes, *The General Theory of Employment, Interest, and Money*, 1936), theoretically and experimentally with a descriptive model of behavior. They bridge the discrepancies of (benchmark) solution concepts and bounded rationality through step-level reasoning, the so-called level-k or cognitive hierarchy models. These models have been recently used as building blocks for new behavioral macro theories to understand puzzles like the lacking rise of inflation after the financial crisis, the effectiveness of quantitative easing, the forward guidance puzzle and the effectiveness of temporary fiscal expansion.

(Published in Special Issue [Bio-psycho-social foundations of macroeconomics](#))

JEL E12 E13 E7 D80 D9 C91

Keywords Beauty contest game; expectation formation; equilibration; level-k reasoning; macroeconomics; game theory; experimental economics

Authors

Felix Mauersberger, University of Bonn, Germany

Rosemarie Nagel, ✉ ICREA-UPF Barcelona GSE, Spain, rosemarie.nagel@upf.edu

Christoph Bühren, TU Clausthal, Germany

Citation Felix Mauersberger, Rosemarie Nagel, and Christoph Bühren (2019). Bounded rationality in Keynesian beauty contests: a lesson for central bankers? Economics Discussion Papers, No 2019-53, Kiel Institute for the World Economy. <http://www.economics-ejournal.org/economics/discussionpapers/2019-53>

1 Introduction

In this paper, we review the results of laboratory experiments on group interaction from a particular angle: we show the importance of those experiments for macroeconomics. In addition, we present the results of an experiment in which the subject group consists of participants in an academic conference on macroeconomics. Based on the experimental results highlighted in this paper, we argue why out-of-equilibrium behavior may have implications for macroeconomics.

Equilibrium is still the key solution concept in macroeconomic analyses. The perhaps most widely used equilibrium concept in macroeconomics is the Rational Expectations Equilibrium (REE). REE presumes that all agents have perfect knowledge about their environment and make on average correct decisions. (Muth, 1961) While REE as a concept remains a useful benchmark and retains macroeconomic analyses tractable, many empirical studies challenge this idea. For instance, there is a large literature that uses survey data on inflation expectations to challenge the validity of the rational expectations assumption. (See Mavroeidis et al. (2014) for a survey.) The notion of rational expectations has also been widely challenged in laboratory studies. (See Hommes (2011) for a survey.)

Due to the critiques regarding the rational expectations hypothesis, we argue in this paper that macroeconomics may benefit from *behavioral* microfoundations. What we consider as behavioral microfoundations are assumptions on individual behavior in an economic model that are well-established in behavioral economics. Since a widely used method in behavioral economics are laboratory experiments, the type of behavioral microfoundation we propose in this paper relies on well-documented findings in controlled laboratory experiments with human subjects. Laboratory experiments have the advantage that they represent controlled environments where human behavior and behavioral responses to exogenous interventions can cleanly be documented.

Specifically, we consider a variation of a game that has been mentioned by Keynes (1936) and has become famous in the behavioral game theory literature: the Beauty Contest (BC) game. Imagine every participant in a group of people is asked to choose a number from 0 to 100. The person whose number is closest to two thirds times the average wins a prize. If all players are rational, the optimal choice in this game is zero, which is the unique Nash equilibrium. This is because if everyone else chooses zero, the choice you need to make is zero. If everyone chooses a different number from zero, a player could deviate and win the game by choosing a lower number. The only number for which that is impossible is the lowest admissible number, zero.

This game has first been introduced into the laboratory by Nagel (1995). The result of this experiment is striking: hardly any subject chooses the Nash equilibrium of zero. Yet, there is a pattern that has lead to the so-called level k reasoning: Instead of finding a fixed point, some subjects choose random strategies, for not understanding the game. If that is assumed by some other players, they will choose $50 \cdot 2/3 = 33.33$. However, someone may argue that then $33.33 \cdot 2/3$ is plausible, etc.. The finding that choices are usually far from zero has been replicated by many experiments both in the laboratory and in the field. Over time there might be gradual convergence to an equilibrium point.

In this paper, we highlight the importance of that game for macroeconomics by arguing that there is a tight link between the BC game and the microfoundations of standard workhorse macroeconomic models. The focus of this paper is on New-Keynesian macroeconomics with standard frameworks as they can be found in textbooks like Woodford, 2003, Galí (2008) and Walsh (2010). Yet, we broach other areas of macroeconomics such as economic growth.

BC games fall in the class of aggregative games. Aggregative games are games in which every player’s payoff depends on the player’s own action and the aggregate of all players’ actions. (Selten, 1970) As shown by authors such as Woodford, 2013 and Angeletos and Lian (2018), agents’ utility (or profit) in New-Keynesian models depends on the agent’s own action and the aggregate of all other agents’ actions in the economy. Specifically, in standard New-Keynesian models, the firms’ optimal price depends on the average price across firms in the economy. Similarly, the household’s optimal consumption depends on the average consumption across households in the economy.

We use the link between BC games and the microfoundations of contemporary macroeconomic models as a motivation to propose the so-called “level k” thinking as a *behavioral* microfoundation for macroeconomics.

Since there has been much critique of the REE and since the evidence of level k as a heuristic of decision-making in the BC game is overwhelming, we argue that considering the implications of level k at the macroeconomic level is a logical step. Since level k is a model of out-of-equilibrium behavior, there is a rationale why macroeconomics may gain insights from considering out-of-equilibrium behavior rather than relying only on equilibrium theory.

The reason why level k is useful for macroeconomics is that it provides inertia in responses to aggregate shocks. García-Schmidt and Woodford (2019) use level k to explain the sluggishness in the response of inflation to low interest rates after the financial crisis 2007-2011. Other applications of level k in macroeconomics include revisiting the effects of forward guidance (Angeletos and Lian, 2018), fiscal expansion Angeletos and Lian (2018), incomplete markets (Farhi and Werning, 2019) and quantitative easing (Iovino and Sergeyev, 2018).

The paper is structured the following way: Section 2 introduces “level k” as an alternative solution concept to rational expectations. Section 3 introduces the Beauty Contest game. In this section, we start from the original framework as it has been long known in the behavioral game theory literature¹ and explain how it generalizes as a canonical framework. Moreover, we review the relevant experimental evidence and compare it to theoretical predictions. Section 4 reviews the relevant microfoundations of macroeconomics and links them with the BC game. Section 5 provides a review of relevant experiments that can be considered as evidence that “level k”-thinking matters at the macroeconomic level. Section 6 concludes.

2 The Level k model

This section reviews the level k model as being introduced into the literature in economics by Nagel (1995). The notion of “level k” reasoning has been shown to provide a realistic description of expectation formation in experimental games, particularly when a game is played for the first time. The basic idea of “level k” reasoning is that, instead of choosing the Nash equilibrium, agents tend to think about what other possibly boundedly rational subjects may choose. The “level k” model of boundedly rational reasoning starts with a specification of a naive approach to the game (“level 0”). Then the model specifies how someone should play who assumes that all other players play naively (“level 1”), how someone should play who assumes that all other players use level 1 reasoning (“level 2”), etc..

The “level k” model is well illustrated using the Beauty Contest game as introduced to economics by Moulin (1986) and first experimented on in Nagel (1995) who at the same time introduced the

¹See Camerer (2011) for a review.

level k model to the literature.² In the standard BC game, agents are asked to simultaneously choose a number between 0 and 100 and the player who has chosen the number closest to two thirds times the average of all players’ numbers wins a fixed prize. The unique Nash equilibrium is that all choose zero. This is because if everyone else chooses zero, the choice someone needs to make is zero. If all players choose a different number from zero, one player could deviate and win the game by choosing a lower number, e.g. two thirds times the other players’ choice. The only number for which that is impossible is the lowest number, being zero.

Let us apply the “level k ” model to this game: Consider the naive player who chooses a random number from 0 to 100 with equal probability. The expected value of this player’s choice is 50. Such behavior is referred to as “level 0.” A slightly more sophisticated player anticipates such behavior and best responds to level 0 by choosing $\frac{2}{3} \cdot 50$. This player type is referred to as “level 1.” Such behavior can also be anticipated and some player may best respond to level 1 and choose $(\frac{2}{3})^2 \cdot 50$. Such a player type is called “level 2.” Similarly, one can define k such thinking steps by “level k ”. A level k player chooses $(\frac{2}{3})^k \cdot 50$. For first period behavior, level zero is typically assumed to be a uniform distribution. In Nagel’s experiment, level 0 in subsequent periods were assumed to be the average of the previous period.

There has been overwhelming evidence that “level k ” has been the heuristic applied by many subjects in experimental games. In many different subject pools, the winner is often a participant who chose a numbers around the level 1 or level 2 response. The experimental evidence on “level k ” is reviewed in section 3.

Following Nagel (1995), the literature has extended the level k model and has addressed possible drawbacks (see the survey of Crawford et al. (2013)). One questionable feature is that the level k model implicitly assumes that *all* other players adopt a certain level of reasoning, assuming that all other choose one level lower than oneself. Stahl and Wilson (1994, 1995) construct level 0, 1, and 2 types as above, and add wordly types which assume that all others are distributed over level 0 and level 1 types. Their experimental set-up (using 3x3 games) does not allow for level 3 or higher types. Camerer et al. (2004) introduce a one parameter cognitive hierarchy model such that level 2 and higher types assume that all other players are distributed according to a Poisson distribution over lower types. Camerer et al. (2004) justify their model by the widely documented finding in the psychology literature that individuals tend to be overconfident.

Alaoui and Penta (2016) outline a model in which players’ depth of reasoning is endogenous. Their motivation is to use the endogenous depth of reasoning (EDR) model to make inferences and sharp predictions that hold across different games. They propose that individuals act as if they face a trade-off between costly reasoning and the benefit of doing so. The costs are related to the game-theoretical sophistication of the player. The benefit instead is related to the game payoffs. Behavior is in turn determined by the individual’s depth of reasoning and his or her belief about the reasoning process of the opponent.

3 Variations of the Keynesian Beauty Contest game

In this section, we introduce some variations of the original Beauty Contest and show to what extent such changes imply differences in the theoretical properties of the new games. Table 1 summarizes the different behavior observed in the BC variations discussed in this section. More abstractly,

²See also, for an historical account, Nagel et al. (2016))

we want to discuss how little changes might or might not change the (rational expectations) equilibrium. As mentioned before, REE is probably the most widely used concept for macroeconomic analyses. However, differences in out-of-equilibrium behavior might also induce behavioral changes. Thus, we want to investigate to what extent variations of game features induce different behavior by humans in laboratory experiments or in the field than one would expect from equilibrium concepts. When behavior differs from equilibrium choices, making predictions using equilibrium concepts is questionable. However, we want to emphasize that this does not make equilibrium concepts disposable. Quite on the contrary, equilibrium solutions can often show how to obtain higher welfare than the welfare subjects actually attain. In such cases, the next step should be to investigate which institutional changes are necessary to attain such desirable outcomes. Also, over time equilibrium can be reached. We will present experimental evidence of the BC variations in this chapter in order to compare behavior to theoretical predictions.

Table 1: Beauty contest experiments from different studies

Authors	BC target	choices	equilibrium	subject pool	no. of subjects*	average	std. dev.	realized target	level 0	level k of target?	spikes at
Nagel (1995)	$p=2/3$	[0,100]	0	students	67	36.73	20.21	24.49	50	k=2	33.33, 22.22
Bosch et al. (2002)	$p=2/3$	[0,100]	0	economists	146	17.15	22.64	11.43		k=4	0
	$p=2/3$	[0,100]	0	newspaper readers	7900	23.08	20.24	15.39		k=3	50, 33.33, 22.22, 0
Nagel et al. (2017)	$p=2/3$, 2 person fixed payoff	[0,100]	0 (dominance)	students (classroom)	160	36.98	22.72	24.65	50	k=2	50, 0
	$p=2/3$, 2 person distance payoff	[0,100]	0	students (classroom)	149	37.36	28.04	24.91	50	k=2	55, 33.33, 0
Benhabib et al.	$p=2/3$	any number	0	students	40	14.03	16.47	9.35	0 ?		no clear spikes (15, 10)
	$p=2/3 + e N(0,10)$	any number	0	students	39	2.37	5.73	1.58			0
Buehren, Nagel (2019)	$p=2/3$, $c=10$	[0,100]	30	students (classroom)	68	42.11	22.35	38.07	50	k=2	50, 43
	$p=2/3$, $c=10$	any number	30	students (classroom)	72	14.53	14.53	19.69	10	k=1 or k=2	10
	$p=4/3$, $c=10$	any number	-30	students (classroom)	17	15.8	16.47	31.07	10		no spikes
	$p=4/3$, $c=-10$	any number	30	students (classroom)	19	18.79	32.36	15.05			no spikes
	$p=1/3$, $c=10$	any number	15	students (classroom)	74	14.56	15.38	14.85	10	k=2	10
	$p=2/3$, $c=10$	[0,100]	30	economists	59	34.98	19.82	33.32	50 (or 10?)	k=4	30
	$p=2/3$, $c=10$	any number	30	economists	65	20.84	15.15	23.89	10	k=3	30
	$p=4/3$, $c=10$	any number	-30	economists	13	31.85	36.06	52.47	10		no spikes
	$p=1/3$, $c=10$	any number	15	economists	11	17.1	12.27	15.7	10		10
Nagel et al. (2017)	$p=2/3$, 2 person fixed payoff	[0,100]	0	students (classroom)	160	36.98	22.72	24.65	50	k=2	50, 0
	$p=2/3$, 2 person distance payoff	[0,100]	0	students (classroom)	149	37.36	28.04	24.91	50	k=2	55, 33.33, 0

3.1 Literature review on the Beauty Contest: Theory and experiments

In this subsection, we exemplify a series of experimental Beauty contest games from the literature, which show features of boundedly rational behavior. The behavioral regularities uncovered in these experiments should be relevant for consideration in behavioral macroeconomic modeling. For more extensive reviews along the same line, refer to Amano et al. (2014), Arifovic and Duffy (2018), Hommes (2013), and Mauersberger and Nagel (2018), which talk about (macro)economic experiments related to the Keynesian Beauty Contest experiments.

The original Beauty Contest: Multiplicity of equilibria The original Beauty Contest from Keynes (1936) constructs a coordination game: “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.” One can interpret that game the following way: We call average preference the “target” to reach. Those not choosing the “average face” earn nothing. In such a game, any face can be an equilibrium choice. That is, if all players choose the same face, then nobody should be able to get a higher payoff by deviating. One can think of a game where the experimenter ranks the faces by a measure of beauty and pay all winners more if they chose a face with a higher rank of the face. Even in that case all faces constitute an equilibrium. Yet, it is not clear whether the most beautiful is picked by all players. In this case, the equilibrium in which all players pick the highest ranked face is called a Pareto-optimal solution. This means that in any other equilibrium payoffs are lower. Such multiplicity is

considered as a serious problem in game theory or other fields, as macro economics. Even a group of rational agents may not know how to coordinate. Indeed, there is high heterogeneity of choices among human subjects, leading to high individual payoffs for some individuals and low payoffs for others (Van Huyck et al., 1990; van Huyck et al., 1991). Therefore, theorists have implemented different ways to select one equilibrium out of the many (see e.g. global game literature, which adds payoff perturbations to obtain a unique equilibrium). We will present another kind of theoretical change.

Variation 1: a unique equilibrium The first experimental study about the basic b-average games ($b=1/2, 2/3, 4/3$) was introduced by Nagel (1995), which contains the first specification and visualization of the so-called level k model, constructed in the same paper. Each subject played within the same group and parameter p for 6 periods. Every subject was only allowed to participate in one session.

When $p < 1$, there is a unique Nash equilibrium, choosing zero, while $p > 1$ has two equilibria, all choose zero, or all choose 100. In a Nash equilibrium, every choice is a best reply, given the other players' choices. An equilibrium of zero exists because if every other player chooses zero, the optimal choice is zero also for $p = \frac{4}{3}$; average is still zero; an equilibrium of 100 exists because if every other player chooses 100, the optimal choice is also 100. The difference between the equilibria is that 0 is an unstable equilibrium, meaning that if one player slightly deviates choices tend not to return to this equilibrium. Consider a deviation from 0 to, say, 1 by one player. Then other players have an incentive to choose a number above 0 and thus shift their behavior away from the equilibrium. Conversely, if one player chooses 99 instead of 100, players still have an incentive to choose 100. Thus, 100 is a stable equilibrium.

We highlight four striking findings. First, in the first period, Nagel finds that zero is rarely chosen and that most choices lie between 20 and 50. Second, she finds that choices are clustered around the numbers that correspond to the levels of reasoning. For the experiment with a target of two thirds times the average, the level k choices correspond to 50 for level 0, $33\frac{1}{3}$ for level 1, $22\frac{2}{9}$ for level 2 etc.. Third, choices tend to get closer to the unique equilibrium as the BC game is repeated. In the version with four thirds times the average, behavior converges to 100, being the stable equilibrium. Fourth, over time, taking the mean of the previous period as level 0, there are only few observations of level k greater than 3 as observed in the first period. Thus, level of reasoning does not increase over time. In a recent working paper, Khaw et al. (2019) investigate to what extent people extrapolate from past experience and to what extent they engage in level-k thinking. They use a similar experimental design to Khaw et al. (2017), in which subjects need to estimate the probability that a green ring is drawn from a virtual ring box. In contrast to Khaw et al. (2017), where this probability is exogenously determined, the probability that subjects need to target is now determined according to $p = c + b \cdot \text{average guess}$. Khaw et al. (2019) do not find much evidence for rational expectations. Instead, the response to changes in the state are inattentive and exhibit much variance. There is much dispersion in the ability of level-0 subjects to track evolution of state. While the strategic sophistication is very heterogeneous across subjects, the largest group of subjects (about one fourth) are the level-1 thinkers.

Variation 2: different subject pools Figure 1 shows the data of over 7,000 subjects playing this game with $b=2/3$, with data of Nagel (1995), and choices from subjects in field studies, including participants from newspaper experiments, classroom and newsgroup and conference data

(see Bosch-Domenech et al. (2002)). As one can see the (average) behavior is far from equilibrium. Yet, there is a pattern (as described in the introduction) across all data sets: the spikes are at 0 (at least in some), 22.22, 33.33 and 50. Only few spikes can be observed at 66.67 and 100. Averages vary from 20 to 35 across all experiments. Thus, the winning number is 2 to 3 steps away from 50, or, in other words, close to level 2 or 3 reasoning, starting at the midpoint 50.

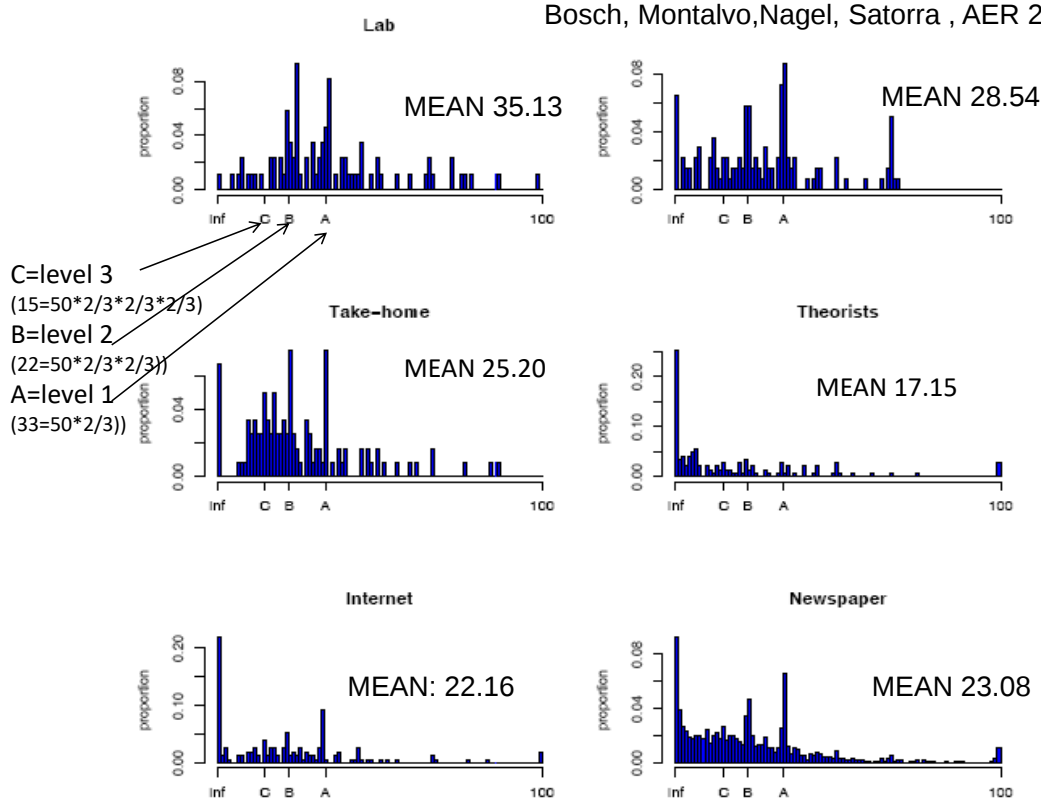


Figure 1: Relative frequencies of choices reported in Bosch-Domenech et al. (2002)

One of the strong conclusions from 30 years of experimental games is indeed that level 2 and 3 are good predictive guesses for best reply behavior, given that there is a (clear) level 0 anchor. In games which include level 1 behavior near an equilibrium, then this point is a good predictor. Games that allow an iterated best reply structure and require many levels of such reasoning are typically far away from the equilibrium, especially in initial periods (see also, e.g. Georganas and Nagel (2011) in toehold auctions).

Variation 3: A unique Pareto optimal solution with distance payoffs Next, we introduce a (small) variation about the payoff structure which has fundamental effects on the theoretical properties also clarifying the importance of benchmark theoretical solutions. We replace the tournament structure (“winner takes it all”) by a distance function. Everyone is paid according to the deviation of their own choice from the target, e.g. the payoff is $100 - (\text{own decision} - \frac{2}{3} \cdot \text{average})^2$. Here, in the Nash equilibrium all players still choose zero. However, now the sum of all payoffs

is maximized, or in other words, a welfare maximizing state is reached in this equilibrium. This is called Pareto optimality, which means that a shift towards the equilibrium makes at least one player better off, without making any other player worse off. Thus, non-compliance will lead to large losses which should be prevented through well-designed institutions. In variation 1, all strategy combinations are Pareto-optimal, as a shift from one winner to another would make the former worse off.

There is another main difference, being particularly interesting for epistemic game theory: iterated elimination of strictly dominated strategies leads to the unique equilibrium. Thus, a rational player who assumes that all others are rational and so on, can come to the conclusion to choose zero. In the prior case, a rational player can only eliminate dominated strategies but not necessarily further domination levels. However, all these theoretical differences have no effect on “boundedly rational” agents. The behavior in the two different games are indistinguishable. (Kocher et al., 2002)

Equilibrium analysis is important, even when it has no predictive power. When behavior fails to reach a welfare optimizing equilibrium. A call for optimal institutional changes is then desirable. As a side note, importance for alterations are also installed when inefficient equilibria are reached as, for example, in public good experiments (see e.g. Gächter et al., 2004). We will see that variation 4, allowing for open intervals, and variation 6 with exogenous signals constitute such rule or institutional changes.

Variation 4: The two players vs many players case (Nagel et al., 2016) Game theory typically starts from the two-player case as a contrast to individual decision-making. Also with many opponents one can form an order statistic like the average or sum to obtain one aggregate decision as in the one-opponent game. The Keynesian BC game constitutes an example of such aggregation to incorporate many players’ behavior into one’s own consideration. The two person $2/3$ average game with fixed payoffs introduces a surprising complication as in a logical puzzle. The reader should pause to think what he would play against an undergraduate student, trying to be closer to the $2/3$ average than the other player to win the prize. It is actually the simplest of all such aggregative games, closely related to the Keynesian Beauty Contest games, yet not even most game theorists see what to do against bounded rational players. The equilibrium is again zero, but with the additional caveat, that it is in (weakly) dominant strategies. This means that it always wins to play zero, but also higher numbers can win against a bounded rational player. The reason is straightforward: If there are two different chosen numbers the lower number always wins, since the midpoint of two numbers multiplied by $2/3$ is always closer to the lower number. When the payoff depends on the above mentioned distance function, then again the iterative elimination procedure leads to zero, yet, with reasoning steps at 100, 50, 25, etc., due to the influence of one’s own number.

However, Figure 2 shows that there are no differences in behavior, neither with two persons vs. many players nor between the two payoff versions (distance vs. tournament payoffs). Dominant strategies constitute the easiest recommendation of what to play in a game. Here, it is even leading to a Pareto-optimal equilibrium. Yet, the equilibrium is hard to find for an inexperienced subject. This is a warning for anybody making predictions, blindly following a desirable theoretical feature. Not even economists (who played prior to seminars of the talk) necessarily pick zero, yet more do so than in the game with more than two players. In the two-player fixed payoff treatment one can test whether a player is rational (chooses a dominant strategy) or not.

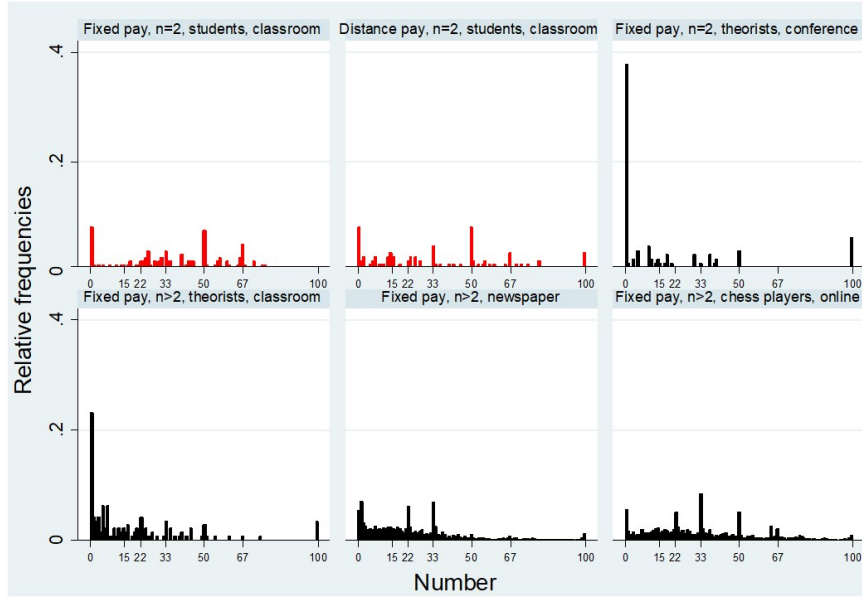


Figure 2: Relative frequencies of choices reported by Nagel et al. (2016) in the two-person games with fixed payoffs (upper left), distance payoffs (upper middle) and with economics professors in various conferences (upper right); lower left for $n > 2$ games with economic professors and students in advanced economic classes; in lower middle newspaper contestants; lower right chess players online. Data source: UR: Grosskopf and Nagel (2008); LL and LM: Bosch-Domenech et al. (2002); LR: Bühren and Frank (2012)

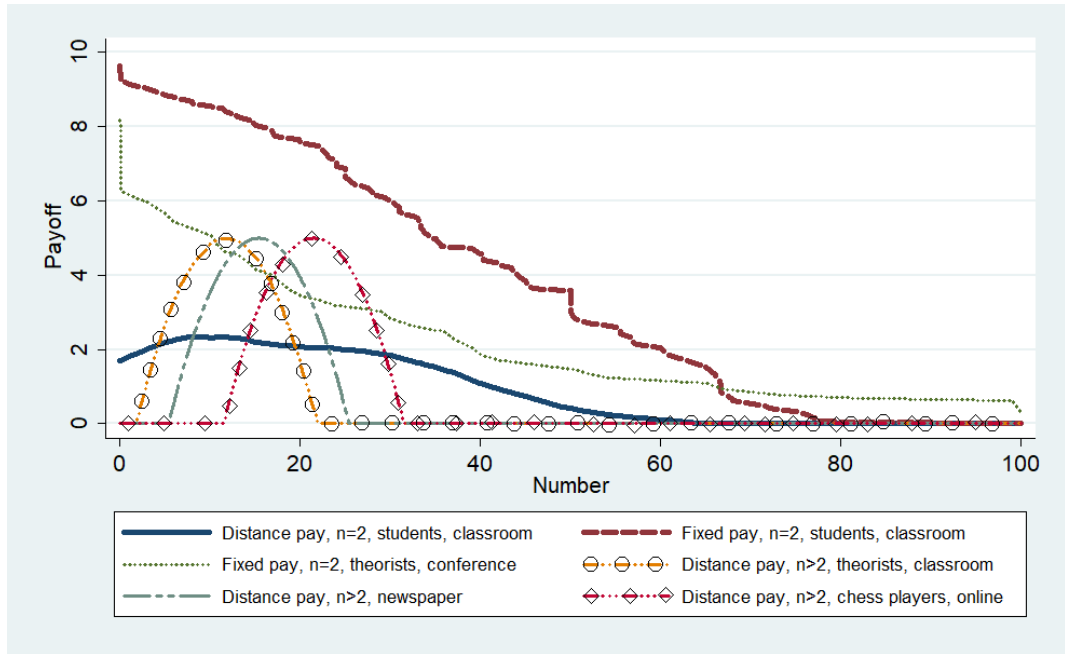


Figure 3: Expected payoffs of choices by treatment (see Nagel et al. (2016))

This is not so easy to see in the games with more than two players as we have seen above, as strictly speaking rationality only precludes numbers above 66.66. The players make several cognitive mistakes in the two-person games. Firstly, they confuse the fixed payoff game with a game to be as close as possible to the target and not just being closer than the other. Furthermore, they do not see the large impact on the target through their own choice. This eludes to the same level k reasoning structure as in the more-than-two-person game, instead of iterating from 50 to 25 etc.. Chou et al. (2009) calls this problem “game form recognition”. In fact, players do not play the game the experimenter suggests to them. In real world situations, we have the problem that parameters and the model itself are not known, not even for the policymaker. If laboratory experiments show such clear discrepancies, then in the real world it might be an even more obvious problem. Figure 3 shows the expected payoff for each choice given the actual distribution in the particular treatment. Obviously, when the weakly dominant equilibrium strategy is played in the fixed payoff treatment, it attains the highest payoff. The optimal outcome in the other games are at or near level 3 or 2, respectively.

Variation 5: Close vs open interval In macroeconomics, it can be very important whether there are boundary restrictions or not in a model. For example, Benhabib et al. (2001) show that if the “zero lower bound” on nominal interest rates is binding, even if the steady state at which monetary policy is active is locally the unique equilibrium, the economy could converge to a steady state in which the nominal interest rate is near zero and inflation is possibly negative.

In the basic BC game, it makes no difference for the equilibrium if there are boundaries or not. If there are no boundaries, any real number can be chosen: positive, negative, or zero. Without boundaries, there is no possibility to formulate an iteration procedure as shown above. An example for a closed choice would be a number from 0 to 100. An example for an open choice set would be any real number. Although the equilibrium is identical, Bühren and Nagel (2019) observe very different behavior in the $2/3$ *average game for both sets. When no boundaries are present, behavior is much closer to zero than in the boundary condition. Level k reasoning is relevant for the closed set. In the open choice set, level 0 can only be zero itself as the midpoint of the real line. Yet, most people seem to assume quite some noise in the behavior of others and thus the average is typically between 10 and 20, albeit far smaller than in the original case, when it is typically above 30 with undergraduate students.

Bühren and Nagel (2019) also compared behavior with and without boundaries in the $2/3$ *average game with an added constant (similar to (Kocher et al., 2002)). When 10 is added to the $2/3$ *average game, the Nash equilibrium is 30, no matter if the choice set is an open interval or between 0 and 100 (see also variation 6 in this section). However, also for the treatments with an added constant, behavior is different with and without boundaries. Bühren and Nagel (2019) show for the no boundaries case that participants typically take 10 as the starting point (anchor), from which some of them proceed with k -level reasoning (level 1 = $2/3 \cdot 10 + 10 = 16.67$, level 2 = 21.11, level 3 = 24.07...). In the boundaries case, participants chose as the starting point mainly 50 (level 1 = $2/3 \cdot 50 + 10 = 43.33$, level 2 = 38.89, level 3 = 35.93...), which led on average to guesses that were again closer to the equilibrium. Therefore in the boundary case the average is 42.11, while, in the unbounded case, the average is 14.52, as shown in table 1.

Variation 6: Strategic substitutes vs compliments So far, we maintained the $\frac{2}{3}$ -average target. In variations 5 and 6, we allow for other targets. In this variation we just change the

sign of the target to $-\frac{2}{3}$ ·average. The equilibrium does not change. Hence, if one uses the Nash equilibrium to make predictions, one would yield the same predictions for the $-\frac{2}{3}$ ·average-game as for the $\frac{2}{3}$ ·average-game. Yet, out-of-equilibrium dramatic changes occur. The reader is invited to see how best replies work (hint: start with any arbitrary average and find the best reply to it and continue such procedure several times). One will quickly see that behavior is closer to equilibrium in the case of a negative sign as compared to a positive sign. In the former, a negative average incurs a positive best reply and vice versa for a positive average, driving the aggregated outcome closer to zero.

The two games differ by their feedback structure. The feedback structure can be characterized by the two related concepts of *strategic complements* and *strategic substitutes*. These two terms were originally shaped by Bulow et al. (1985) and developed for studies of firm interaction but can be applied to any game or situation of strategic interaction between different agents. Different names under which these terms are known in the literature but which generally describe the same idea are “negative feedback” (strategic substitutes) and “positive feedback” (strategic complements.)

Agents’ decisions are complements if they have an incentive to match other agents’ decisions. Conversely, agents’ decisions are substitutes if agents have an incentive to do the opposite of what others are doing. For instance, if a firm can increase its profit by charging the same price as others, then prices are strategic complements. If firms can make more profit by charging high prices when their competitors charge low prices (or vice versa), then prices are strategic substitutes.

In simple BC games where agents need to be closest to a target of b times the average, whether the system exhibits strategic substitutes or strategic complements only depends on the coefficient b . If $b > 0$, the system exhibits strategic complements, since if all others increase their choices, an individual also has an incentive to increase her choice. Conversely, if $b < 0$, the system exhibits strategic substitutes, since if all others increase their choices, the individual has an incentive to decrease her choice.

The basic BC games are ideal to show the difference between these theoretical structures. As seen in Figure 5, strategic substitutes lead to much closer equilibrium choices than in a situation with strategic complements. An arbitrarily high average belief in the latter will induce a high best reply, while in the former a low negative choice will result. Thus, if people just differ by one reasoning step, the average of such types will come closer to equilibrium in strategic substitute cases.

Assenza et al. (2018) give an example of a policy recommendation that where an environment with strategic complements changes to an environment with a mixture of strategic complements and strategic substitutes. They show that disobeying Taylor principle in a New-Keynesian framework (i.e. $\phi_\pi > 1$; see section 4.2.3) leads to a system which purely exhibits strategic complements. However, fulfilling the Taylor principle introduces strategic substitutability into the system so that the New-Keynesian framework becomes a mixture between strategic substitutes and strategic complements. Their experimental results show that if the Taylor principle is not fulfilled, outcomes generally do not converge to the RE steady state. However, they show in their experiments that just obeying the Taylor principle ($\phi_\pi = 1.005$) does not result into convergence. Yet, their experimental results suggest that a stronger reaction coefficient ($\phi_\pi = 1.5$) is sufficient to ensure convergence to the steady state.

Variation 7: Signals with (non)zero means (Benhabib et al., 2019) After about 20 years of BC experiments, a well-documented experimental result is that initial behavior is far from

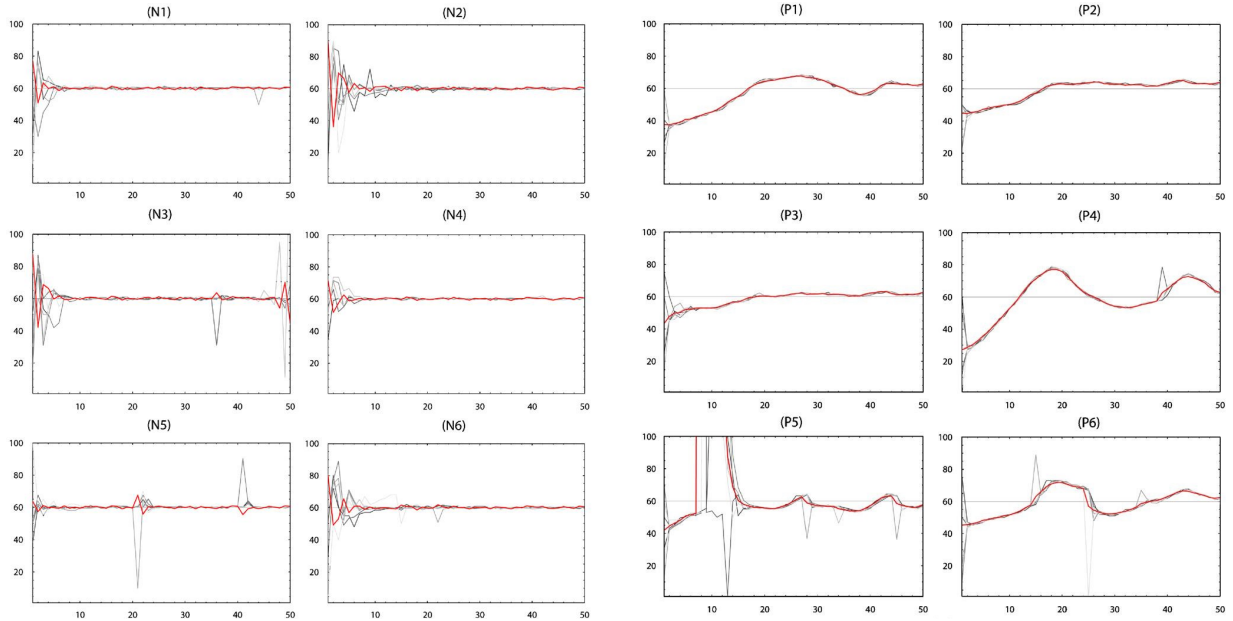


Figure 4: Experiments by Heemeijer et al. (2009) with strategic substitutes (left); experiments with strategic complements (right)

equilibrium and sluggishly unravels over time to the equilibrium. However, this sluggish adjustment can be very costly, especially for variations with distant payoff structures as stated in variation 2. This provides a rationale for academicians and policymakers to develop institutions to offset such deviations.

Perhaps the most useful variation so far has been where every subject i receives an individual idiosyncratic signal, which is added to the original target: $2/3 \cdot \text{average} + \epsilon_i$. Subjects know the distribution of the signals but are not informed about the signals which other subjects receive. ϵ_i is drawn from a normal distribution with mean zero and a positive variance. Choices therefore need to be made from the entire real numbers (open interval) as introduced in variation 4.

An equilibrium choice is that every player chooses her own signal. While computing the equilibrium is more elaborate than in all other cases, the intuition why choosing one's own signal is an equilibrium is simple: if all players choose their own signal, the mean will be zero in expectation. Thus, $2/3$ -average is zero and the target becomes each player's own signal ϵ_i . Thus, if every player chooses her own signal ϵ_i , every player at the same time chooses her target. Finally, the intuitive answer is that the "anchor" ϵ_i should be chosen. Indeed, on average, subjects choose such an anchor.

Yet, it is too early to be content. From an experimentalist's perspective, signals have to be chosen wisely. However, what does that mean in a real environment?

When the mean of the normal distribution is not zero, then ϵ_i is not an equilibrium choice. For example, if the mean of the distribution is 10, then the equilibrium choice of an individual subject i is $20 + \epsilon_i$. Thus, in equilibrium subjects choose 30 on average. Here again subjects fail to make the right equilibrium choice. (Benhabib et al., 2019)

4 Microfoundations of macroeconomics

Mauersberger and Nagel (2018) show that many games and models can be considered as Beauty Contest games. Among them are Cournot competition, Bertrand competition, auctions, asset markets, Cobweb models, ultimatum games, stag-hunt games, New-Keynesian models, the 11-20 game and global games. This section explains in some detail why standard workhorse models in macroeconomics can be considered as a BC game.

Section 4.1 introduces the generalized beauty contest game as discussed by Mauersberger and Nagel (2018). Section 4.2 establishes the link between the standard New-Keynesian model and aggregative games. We first consider the standard New-Keynesian model in section 4.2.1, in which the relationships between the macroeconomic variables have been derived under the assumption of a rational, representative agent. Under certain restrictions on the expectations operator, expectations in those first-order conditions can be replaced by boundedly rational forecasts. We do not explain the details of these restriction in this chapter and refer the interested reader to Branch and McGough (2009). In section 4.2.2, we consider a more microfounded version of the New-Keynesian model that allows for quite arbitrary boundedly non-rational individual forecasts. Section 4.3 establishes the link between growth models and BC games. To the best of our knowledge, this latter link has previously not been established yet.

4.1 Generalization of the Beauty Contest game

Now consider a generalization of the BC game in order to present a clear relation between experimental BC games and the microfoundations of macroeconomics. The BC game in which players have to reach a target that involves the average falls in the class of *aggregative games*. Aggregative games have been introduced by Selten (1970). Aggregative games are defined as games in which the payoffs of every player are a function of their own strategy and the sum of all players' strategies. Since the mean is the sum divided by the number of players, BC games with a target containing the average fall into the class of aggregative games.

Aggregative games have been generalized to so-called *fully aggregative games* (Cornes and Hartley, 2012). Cornes and Hartley (2012) propose a more general aggregation function. They require that the aggregation function $g(\cdot)$ needs to be additively separable, i.e. for the strategies of all players $y_1, \dots, y_N$, there exist increasing functions $h_0, h_1, \dots, h_N : \mathbb{R} \rightarrow \mathbb{R}$ such that $g(y_1, \dots, y_N) = h_0(\sum_{i=1}^N h_i(y_i))$. It is easy to see that any aggregative game is also a *fully aggregative game*. Acemoglu and Jensen (2013) analyze how a change in the exogenous parameters affects the Nash equilibrium, also considering the cases of strategic substitutes.

We define the notion of *generalized Beauty Contest game* following Mauersberger and Nagel (2018). While we consider the one-dimensional case where players choose among real numbers, we consider the multidimensional case (and an example thereof) in the next section 4. We refer to a game as a *generalized Beauty Contest game* if the best response of a player i at time t can be written as

$$y_t^i = \hat{E}_t^i \{ c + b \cdot f(y_{1,t}^1, y_{1,t}^2, \dots, y_{1,t}^N) + d \cdot f(y_{t+1}^1, y_{t+1}^2, \dots, y_{t+1}^N) + \epsilon_t^i \} \quad (1)$$

where i is the individual-superscript, N is the number of individuals, c, b, d are coefficients and the function $f(\cdot)$ is the aggregation across all individuals. We allow for several different aggregations. $f(\cdot)$ can be the average, the median, the sum, the minimum, the maximum, the mode, the action least chosen or the action chosen by at least $h \leq N$ individuals. ϵ_t^i is an idiosyncratic shock.

Thus, if the function $f(\cdot)$ corresponds to the sum or the average, the game falls into the class of *aggregative games* and in the class of *fully aggregative games*. If the function $f(\cdot)$ corresponds to other order statistics (minimum, maximum, mode, the action least chosen or the action chosen by at least $h \leq N$ individuals), then the game does not fall into the class of *fully aggregative games*, since these are not additively separable functions.

Figure 6 shows the relative frequency of behavior in such a case. Finally, behavior closely corresponds to the equilibrium with a very small variance, contrasted to figure 4, without signals but also with an open interval. The naive or intuitive choice, “choose your signal” (as an anchor) is chosen by the majority of subjects (see Benhabib et al. (2019)).

When idiosyncratic signals are drawn from a normal distribution with positive mean (say 10), then behavior again is far away from equilibrium when high level-ks are necessary to reach such an equilibrium. The reason is that human subjects again use there signal or close to it as their choice. As a consequence, average behavior is close to the average signal, 10 in our example. However, $2/3 \times 10 + \text{expected signal (10)}$ is 16.67. Further iterations slowly leads to 30, the average equilibrium of such a game (see also variation 4). Thus, signals have to be chosen wisely (Bühren and Nagel, 2019).

There is of course, one caveat. What do such signals mean in reality, or how can a policy maker choose or distribute such signals. Or are they ”constructed” by the agents, under concern, themselves.

4.2 New-Keynesian models

4.2.1 Standard textbook New-Keynesian model

A heterogeneous expectations version of the textbook New-Keynesian model as encountered in Woodford (2003), Galí (2008) or Walsh (2010) can, under some restrictions of the expectations operator (see Branch and McGough (2009)), be written as

$$y_t = \bar{y}_{t+1}^e - \sigma(i_t - \bar{\pi}_{t+1}^e - \rho) \quad (2)$$

$$\pi_t = \kappa y_t + \beta \bar{\pi}_{t+1}^e \quad (3)$$

$$i_t = \rho + \phi_\pi(\pi_t - \pi) \quad (4)$$

where y_t denotes the output gap, π_t inflation, t the time-subscript and $\bar{\pi}_{t+1}^e$ and $\bar{y}_{t+1}^e$ the mean expected future values of output gap and inflation, being the average forecast of all subjects. The model is closed under an inflation targeting rule for the nominal interest rate i_t with a constant inflation target π . It is easy to see that π is the unique rational expectations equilibrium.

We consider a learning-to-forecast game, in which subjects are only paid for forecasting inflation and output gap. After everyone submitted their forecasts, actual output gap and inflation are generated by equations (2) and (3). Note that inflation and output gap of period t depend on $\bar{\pi}_{t+1}^e$ and $\bar{y}_{t+1}^e$, respectively, which are the means across all agents’ forecasts for period $t+1$. An individual subject i is then paid according to a distance function such as $U_{t+1}^i = A - Q(\hat{E}_t^i \pi_{t+1} - \pi_{t+1})^2$. A and Q are positive constants. This distance function measures how close the agent’s inflation forecast for period $t+1$ (given in period t), $\hat{E}_t^i \pi_{t+1}$, is to the actual inflation in $t+1$, π_{t+1} . This is analogous for the output gap forecast.

For the sake of simplicity, we describe the New-Keynesian model as a univariate forecasting game of inflation at time t , π_t .³

To reduce the dimensionality of this forecasting game to a single dimension, we use the ad-hoc assumption that expectations of the output gap are equal to its long-run steady state value, obtained by using equation (3):

$$\bar{y}_{t+1}^e = \kappa^{-1}(1 - \beta)\pi \quad (5)$$

Substitute (5) into (2) to obtain

$$y_t = \kappa^{-1}(1 - \beta)\pi + \sigma\phi_\pi\pi - \sigma\phi_\pi\pi_t - \sigma\bar{\pi}_{t+1}^e \quad (6)$$

$$\pi_t = \kappa y_t + \beta\bar{\pi}_{t+1}^e \quad (7)$$

Insert (6) into (7) one yields the value of inflation that subjects needed to forecast at time $t - 1$:

$$\pi_t = c + d\bar{\pi}_{t+1}^e \quad (8)$$

with $c \equiv \frac{(1-\beta)+\kappa\sigma\phi_\pi}{1+\kappa\sigma\phi_\pi}\pi$ and $d \equiv \frac{\beta-\kappa\sigma}{1+\kappa\sigma\phi_\pi}$. It becomes apparent that the New-Keynesian model can be considered a Beauty Contest game with $c > 0, d > 0, b = 0, f(\cdot) = \bar{\pi}_{t+1}^e$ being the average. A learning-to-forecast experiment, in which subjects' only task is forecasting inflation, based on the model equation of the New-Keynesian framework (8) has first been introduced by Pfajfar and Žakelj (2014). These authors vary the Taylor rule for different treatments, and find that the variance in inflation expectations decreases starkly as the computerized central bank adopts a more aggressive response to deviations of inflation from the target.

4.2.2 A more microfounded New-Keynesian model

Woodford (2013) and García-Schmidt and Woodford (2019) consider a version of the New-Keynesian model that is populated by heterogeneous households and firms that do not necessarily have rational expectations. Specifically, these papers assume that forecasts are not necessarily model consistent, not necessarily the same across agents and expectations are not necessarily those that agents previously expected to hold.

In their framework, it is shown that the optimal choice of consumption and price not only depend on the macroeconomic outcomes in the next period but instead on the outcomes far into the future:

$$c_t^i = \sum_{k=0}^{\infty} \beta^k \hat{E}_t^i \{ \lambda c_{t+k} - \sigma(i_t - \pi_{t+k}) \} \quad (9)$$

$$\pi_t^j = \sum_{k=0}^{\infty} (\alpha\beta)^k \hat{E}_t^j \{ \phi c_{t+k} + \xi \pi_{t+k} \} \quad (10)$$

where t is the time-period subscript, i is the household superscript, j is the firm superscript, and $\hat{E}_t^i$ denotes the boundedly rational expectation held at time t . β is the discount factor that discounts

³One can also consider this model as a BC game, in which the variable of interest to be a vector. Anufriev et al. (2019) conduct an experiment on this type of game.

future payoffs and α is an exogenous probability that the price remains the same as in the last period $t - 1$. $\lambda, \sigma, \phi, \xi$ are positive constants.

Equation (9) presents the consumption rule for the households: c_t^i is a measure of individual total consumption expenditure at time t , c_t denotes average consumption expenditure at time t . i_t is the nominal interest rate which is set by the central bank and can thus be arbitrarily specified. Thus, the (rational expectations) equilibrium and whether it is unique totally depends on the monetary policy rule. Equation (10) presents the pricing rule of the firms: π_t^j denotes firm j 's individual inflation at time t (compared to average prices at time $t - 1$) and π_{t+k} denotes average inflation at time t .

Equations (9) and (10) have a recursive form so that they can be rewritten as

$$c_t^i = \lambda \hat{E}_t^i c_t - \sigma(i_t - \hat{E}_t^i \pi_t) + \beta \hat{E}_t^i c_{t+1}^i \quad (11)$$

$$\pi_t^j = \phi \hat{E}_t^j c_t + \xi \hat{E}_t^j \pi_t + \alpha \beta \hat{E}_t^j \pi_{t+1}^j \quad (12)$$

Thus, individually optimal consumption and prices depend on agents' beliefs about the average consumption and price in the economy and the forecast of agents' own optimal choice in the next period. This equation can be written in matrix form in the following way:

$$\begin{bmatrix} c_t^i \\ \pi_t^j \end{bmatrix} = \begin{bmatrix} -\sigma i_t \\ 0 \end{bmatrix} + \begin{bmatrix} \lambda & \sigma \\ \phi \xi & \end{bmatrix} \begin{bmatrix} \hat{E}_t^i c_t \\ \hat{E}_t^j \pi_t \end{bmatrix} + \begin{bmatrix} \beta & 0 \\ 0 & \alpha \beta \end{bmatrix} \begin{bmatrix} \hat{E}_t^i c_{t+1}^i \\ \hat{E}_t^j \pi_{t+1}^j \end{bmatrix} \quad (13)$$

Strictly speaking, equation (13) does not fit into our scheme of a BC game in equation 1 because of the terms $\hat{E}_t^i c_{t+1}^i$ and $\hat{E}_t^j \pi_{t+1}^j$. At the same time, (13) provides an interesting game structure that has been analyzed as a laboratory experiment by Mauersberger (2016).

However, by adopting a simplified view, one could consider the New-Keynesian model in (13) as a BC game by crudely considering $\begin{bmatrix} \beta & 0 \\ 0 & \alpha \beta \end{bmatrix} \begin{bmatrix} \hat{E}_t^i c_{t+1}^i \\ \hat{E}_t^j \pi_{t+1}^j \end{bmatrix}$ as an idiosyncratic term ϵ_t^i - at least from the analysts perspective. Thus, a paper that can be considered as a (simplified) experimental implementation of equation (13) is Benhabib et al. (2019). The differences between Benhabib et al. (2019) and Mauersberger (2016) are: (a) that Benhabib et al. is a static BC game; (b) most importantly, that Benhabib et al. specify $\hat{E}_t^i c_{t+1}^i$ (or $\hat{E}_t^j \pi_{t+1}^j$) as an exogenous shock while in Mauersberger (2018), $\hat{E}_t^i c_{t+1}^i$ and $\hat{E}_t^j \pi_{t+1}^j$ are forecasts and thus an choices of the subjects.

4.2.3 Strategic substitutes and strategic complements

A useful concept for understanding the influence of aggregate behavior on individual decisions is the distinction between “strategic substitutability” and “strategic complementarity.” The terms were originally shaped by Bulow et al. (1985) and developed for studies of firm interaction but can be applied to any game or situation of strategic interaction between different agents. Different names under which these terms are known in the literature but which generally describe the same idea are “negative feedback” (strategic substitutes) and “positive feedback” (strategic complements.)

Agents' decisions are complements if they have an incentive to match other agents' decisions. Conversely, agents' decisions are substitutes if agents have an incentive to do the opposite of what others are doing. For instance, if a firm can increase its profit by charging the same price as others, then prices are strategic complements. If firms can make more profit by charging high prices when other their competitors charge low prices (or vice versa), then prices are strategic substitutes.

In simple BC games where agents need to be closest to a target of p times the average, whether the system exhibits strategic substitutes or strategic complements only depends on the coefficient p . If $p > 0$, the system exhibits strategic complements, since if all others increase their choices, an individual also has an incentive to increase her choice. Conversely, if $p < 0$, the system exhibits strategic substitutes, since if all others increase their choices, the individual has an incentive to decrease her choice.

The New-Keynesian model is a more complicated environment than simple BC games. In the New-Keynesian model, there are two endogenous variables, inflation and consumption, and the interest rate, a policy variable. Depending on the policy reaction to inflation volatility, the system may exhibit purely strategic complements or a mixture of strategic substitutes and complements. See Assenza et al. (2018) for a more detailed exposition.

Equations (11) and (12) show that the only source introducing strategic substitution into the system is the policy interest rate of the central bank. A commonly used interest rate rule is

$$i_t = \bar{i} + \phi_\pi(\pi_t - \pi) \quad (14)$$

where $\bar{i}$ is a constant corresponding to the long-run steady state of the interest rate and π is the central bank's inflation target. The coefficient ϕ_π captures the strength of the response of the central bank to inflation fluctuations. Using (14) in (11) shows that the monetary policy authority introduces some strategic substitutability into the system if $\phi_\pi > 1$. Then the New-Keynesian model becomes a mix between strategic substitutes and strategic complements.

The condition $\phi_\pi > 1$ is also the sufficient conditions that guarantees a unique rational expectations equilibrium. In the absence of any shock, the rational expectations equilibrium becomes π for inflation and the long-run steady state $\frac{\sigma(\bar{i}-\pi)}{\lambda+\beta-1}$ for consumption.

4.3 Growth

This section is a digression to long-run macroeconomics. By *long-run*, we mean that the capital stock is not fixed, being a key assumption in neoclassical growth models. Standard growth models like the Solow-Swan model (Solow, 1956) and the Ramsey-Cass-Koopman models (Ramsey, 1928; Cass, 1965; Koopmans, 1965) describe the growth of the capital stock in the following way:

$$K_{t+1} - K_t = (f(K_t) - c_t) - \delta K_t \quad (15)$$

K_t is the aggregate capital stock of the economy in any period t , f is the neoclassical production function, c_t the consumption and delta the depreciation rate.

Lei and Noussair (2002) implement this model as a coordination game with a decentralized market for capital. In this version, the economy is populated by a finite number of individuals N , which are indexed by j . The capital stock of the economy is then the sum of individual capital holdings k_t^j : $K_t = \sum_{j=1}^N k_t^j$. With e.g. a quadratic utility function $U(c_t) = 310C_t - 5C_t^2$, there exist a unique pair (C^*, K^*) of optimal consumption and capital that determine the so-called golden rule steady state. Social welfare is thus maximized if

$$K_t = \sum_{j=1}^N k_t^j = K^* \quad (16)$$

which can, for a specific individual i , be rewritten as $k_t = K^* - \sum_{j=1, j \neq i}^N k_t^j$. This is a BC game with $c = K^*$ and $b = -1$ and f being the sum of the investment choices of all individuals.

5 Behavioral microfoundation of macroeconomics

There is a recent growing literature, exploring the implications of finite depths of reasoning in macroeconomics.⁴ One technical difference is that the microeconomic literature on Beauty Contest games rather deals with *static* BC games, closely related to Keynes newspaper contests, i.e. setups in which agents need to form beliefs about the current decisions of others. Contrary to that, papers in macroeconomics also study *dynamic* BC games, in which agents must form beliefs about future actions of others.

Early approaches of level k in macroeconomics include Evans and Ramey (1992), who introduce calculation costs for using the (correct) model equations. Another early approach introduced by Guesnerie (1992, 2009) “eductive stability” of the REE. The question of eductive stability is whether for any initial (naive) belief, outcomes converge to the REE using infinitely many iterations of the model equations.

More recently, García-Schmidt and Woodford (2019) introduce level k into a New-Keynesian model that allows for heterogeneous boundedly rational forecasts in the same way as Woodford (2013). The motivation for their paper is the empirical observation that a prolonged period of low nominal interest rates during the financial crisis has not resulted in high inflation. This has led to increased interest in the “Neo-Fisherian” hypothesis according to which low nominal interest rates may themselves cause lower inflation. García-Schmidt and Woodford (2019) challenge the Neo-Fisherian paradox by proposing that agents in their model use level k to form expectations. Under level k , a commitment to maintain a low nominal interest rate for longer should always result in both higher output and higher inflation rather than deflation. However, in the case of a long-horizon commitment to low interest rates, the expansionary or inflationary effects are less pronounced than under rational expectations without any uncertainty. The explanation for this is as follows: They assume that the expectations of the naive level-0 type correspond to a state of the economy in which no shock has occurred. Since other types respond to such behavior, there will be no deflationary spiral. However, the response to the naive type and other boundedly rational types causes inertia in the adjustment. In other words, there will not an immediate adjustment to high output and high inflation as in the rational case.

Angeletos and Lian (2018) study the effects of uncertainty about others’ actions in a large class of games that nests but is not limited to the New-Keynesian model. Based on the insight that the New-Keynesian model can be considered as a BC game, Angeletos and Lian obtain the following results in the absence of common knowledge and the presence of higher-order beliefs (with a level of reasoning greater than one): first, any general-equilibrium effect is mitigated for a coefficient smaller than one. Provided that the coefficient is smaller than one, all things equal, agents with higher-order beliefs are less responsive than agents with lower-order beliefs. This is because after the shock, the level-0 type adjusts his behavior immediately and expects to optimize choosing the previous period average. The higher the level of reasoning the further away is an agent from the level-0 choice and the closer she is to the rational expectations equilibrium. Second, the further in the future an event occurs, the smaller the effect on the present - a phenomenon that they call “horizon effect”. This can be explained by the fact that longer horizons involve more iterations of the forward-looking, Euler-type equations, which have a dampening effect for beliefs of higher

⁴Alternative approaches to the level- k literature to introduce bounded rationality into macroeconomics have been adaptive learning (see e.g. Evans and Honkapohja (2001)), diagnostic expectations (Bordalo et al., 2018), limited attention (see e.g. Sims, 2003, Maćkowiak and Wiederholt (2009) and Gabaix (2014))

order. Third, under quite general conditions, the effect size goes to zero as the horizon $T \rightarrow \infty$. This is due to the fact that infinite levels of reasoning are anchored to the common prior and are hence unresponsive to the news even in the presence of small idiosyncratic shocks.

Angeletos and Lian (2018) show that this has two important implications for New-Keynesian macroeconomics: first, it solves the forward-guidance puzzle. The forward guidance puzzle is the finding that, under common knowledge, a credible announcement regarding the monetary policy of e.g. 1,000,000 years in the future has the same or even greater effects today as an announcement regarding the monetary policy of tomorrow. Second, under common knowledge, it has been shown that the current effects of a fiscal expansion of a given magnitude increases as the stimulus is announced to take place further in the future. This is no longer the case under imperfect knowledge, which provides a rationale for “front-loading.”

Farhi and Werning (2019) present another application of level k in macroeconomics. They show their results both qualitatively and quantitatively. They demonstrate that future interest rate changes have smaller effects than present interest rate changes under level k thinking, which they call “mitigation effect.” Like Angeletos and Lian (2018), they also find that introducing level- k thinking leads to a “horizon effect”, i.e. interest rate changes further out in the future have smaller effects in the presence. However, their calculations show that both mitigation and horizon effect under level- k thinking are relatively modest in size. They show that a sizable effect is only yielded by combining level k with incomplete markets, introducing uninsurable idiosyncratic risk.

Iovino and Sergeyev (2018) investigate the implications of standard central bank balance sheet policies, i.e. quantitative easing and foreign exchange interventions, under the assumption that agents engage in level- k thinking in an overlapping generations model. They find different implications from the rational expectations case: First, in contrast to rational expectations, where central bank implications are neutral, central bank interventions are effective under level- k thinking under mild conditions. The reason is that since agents do not hold rational expectations about future endogenous variables, they underestimate the tax risk resulting from policy interventions and incorrectly forecast future assets prices. As a result, they demand lower risk premia, which increases asset prices and thus renders balance sheet policies effective. Second, they show that individual and cross-sectional average forecast errors about future endogenous variables are predictable after balance sheet interventions. They validate their predictions, using data on the mortgage purchases by US enterprises as a proxy for quantitative easing.

Recently, the idea of higher-order beliefs in forecasting has also raised interest in the empirical literature in macroeconomics. Coibion et al. (2018) introduce a survey to New Zealand firm managers, not only eliciting their expectations about inflation but also asking them about their beliefs of other managers and measuring their depth of reasoning in an incentivized p-beauty-contest task. They also investigate whether managers’ beliefs influence their actions, finding that firms obtaining any type of information made significant reductions in employment and investment but not in prices or wages. Furthermore, the authors demonstrate differences to standard implementations of k -level modeling. First, most managers that exhibit a level of reasoning lower than 4 believe others will submit an answer in the same range as theirs. Thus, they do not select a number equal to two-thirds times the estimated average. Second, managers generally believe that some of the other managers exhibit higher levels of reasoning than they do. In contrast to that, level k and other models assume that agents act as if all other agents are lower-level thinkers than them.⁵ Third,

⁵Stahl (1993) in a theoretical model introduces an abound number of thinking types, from those who think that others are below them and others are above them, with different informational assumptions about other players

managers highly underestimate the dispersion of the answers. Altogether, they find no significant evidence that agents' degree of k -level thinking is related to either how they update their beliefs based on new information or how changes in information affect their decisions.

6 Conclusion

By reviewing the literature on Beauty Contest games, we hope to have convinced the critical reader that using level k as a behavioral microfoundation for macroeconomic modeling is a promising approach. There is a lot of empirical evidence that level k is a widely used heuristic in experimental games. Level k allows for iterated best response structures, similarly as in the BC games. In section 4, we explain the link between macroeconomic models and experimental BC games. The first-order conditions of the optimization problem of the household and the firm in New-Keynesian models can in fact be considered as BC games. Since level k has been argued to be a good description of how experimental subjects in BC games make their choices, the logical next step is to analyse such implications for macroeconomics. We have reviewed the pioneer work on that in section 5.

The virtue of level k is that it specifies actions, typically different for naive vs. more sophisticated players, based on cognitive reasoning processes (see e.g. Marchiori et al., 2019) who develop a set of heuristics build on level k). Thus, building level k into economic analyses can relatively easily be done in a way to retain these analyses tractable. However, we have also discussed the limitations and drawbacks of level k in section 2. Hence, level k is certainly not the end of the growing field of behavioral macroeconomics.

Yet, we also showed several virtues of the concept of the rational expectation equilibrium and off equilibrium structures as iterated best reply or iterated elimination of dominated strategies. First of all, behavior may converge in the long run to the (static) equilibrium, even when players do not know the structure of the model (see e.g. Assenza et al., 2018). Second, and equally important, even if the equilibrium is unique, Pareto optimal and welfare maximizing choices may not correspond to equilibrium initially or even not over time. Boundedly rational agents might never converge to equilibrium play due to cognitive difficulties. In such cases, there is a call for developing ideas how to induce convergence instantaneously, as demonstrated in our last variation of section 3.1 with idiosyncratic shocks. Those signals can serve as anchors and thus can even induce naive players to play the equilibrium. When equilibrium play is not desirable, as it is the worst possible outcome, e.g., in a public goods game, then policy should shift behavior away from such an equilibrium, for example, with punishment in public goods games.

References

- Daron Acemoglu and Martin Kaae Jensen. Aggregate comparative statics. *Games and Economic Behavior*, 81(1):27–49, 2013.
- Larbi Alaoui and Antonio Penta. Endogenous Depth of Reasoning. *Review of Economic Studies*, 83(4):1297–1333, 2016.
- Robert Amano, Oleksiy Kryvtsov, and Luba Petersen. Recent Developments in Experimental Macroeconomics. *Bank of Canada Review*, pages 1–11, 2014.

distributions.

- George Marios Angeletos and Chen Lian. Forward guidance without common knowledge. *American Economic Review*, 108(9):2477–2512, 2018.
- Mikhail Anufriev, John Duffy, and Valentyn Panchenko. Planar Beauty Contests. *SSRN Working Paper*, 2019.
- Jasmina Arifovic and John Duffy. Heterogeneous Agent Modeling: Experimental Evidence. In Cars H. Hommes and Blake LeBaron, editors, *Handbook of Computational Economics, Vol. 4*, chapter 9, pages 491–540. North-Holland, Amsterdam and Oxford, 2018.
- T Assenza, P Heemeijer, C H Hommes, and D Massaro. Managing Self-organization of Expectations through Monetary Policy : a Macro Experiment. *TSE Working Paper*, 2018.
- Jess Benhabib, Stephanie Schmitt-Grohé, and Martín Uribe. The Perils of Taylor Rules. *Journal of Economic Theory*, 96:40–69, 2001.
- Jess Benhabib, John Duffy, and Rosemarie Nagel. De-framing rules to De-anchor Beliefs in Beauty Contest Experiments: Keynesian level-k vs. Keynesian Sentiments. *mimeo*, 2019.
- Pedro Bordalo, Nicola Gennaioli, and Andrei Shleifer. Diagnostic Expectations and Credit Cycles. *The Journal of Finance*, 73(1):199–227, feb 2018.
- Antoni Bosch-Domenech, Jose G Montalvo, Rosemarie Nagel, and Albert Satorra. One, Two(Three), Infinity,..., : Newspaper and Lab Beauty-Contest Experiments. *The American Economic Review*, 92(5):1687–1701, 2002.
- William A. Branch and Bruce McGough. A New Keynesian model with heterogeneous expectations. *Journal of Economic Dynamics and Control*, 33(5):1036–1051, 2009.
- Christoph Bühren and Björn Frank. Chess Players Performance Beyond 64 Squares: A Case Study on the Limitations of Cognitive Abilities Transfer. *Talent Development and Excellence*, 4(2): 157–169, 2012.
- Christoph Bühren and Rosemarie Nagel. Strategic Nudging in the Beauty Contest Game. *mimeo2*, 2019.
- Jeremy I. Bulow, John D. Geanakoplos, and Paul D. Klemperer. Multimarket Oligopoly: Strategic Substitutes and Complements. *Journal of Political Economy*, 93(3):488–511, 1985.
- Colin F. Camerer. *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, Princeton, NJ, 2011.
- Colin F Camerer, Teck-hua Ho, and Juin-Kuan Chong. A Cognitive Hierarchy Model. *Quarterly Journal of Economics*, 119(3):861–898, 2004.
- David Cass. Optimum Growth in an Aggregative Model of Capital Accumulation. *The Review of Economic Studies*, 32(3):233, jul 1965.
- Eileen Chou, Margaret McConnell, Rosemarie Nagel, and Charles R. Plott. 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, 2009.

- Olivier Coibion, Yuriy Gorodnichenko, Saten Kumar, and Jane Ryngaert. Do You Know That I Know That You Know...? Higher-Order Beliefs in Survey Data. *NBER Working Paper Series*, 2018.
- Richard Cornes and Roger Hartley. Fully aggregative games. *Economics Letters*, 116(3):631–633, 2012.
- Vincent P. Crawford, Miguel A. Costa-Gomes, and Nagore Iriberri. Structural models of nonequilibrium strategic thinking: Theory, evidence, and applications. *Journal of Economic Literature*, 51(1):5–62, 2013.
- George W. Evans and Seppo Honkapohja. *Learning and Expectations in Macroeconomics*. Princeton University Press, Princeton, jan 2001. ISBN 9781400824267.
- George W Evans and Garey Ramey. Expectation Calculation and Macroeconomic Dynamics. *American Economic Review*, 82(1):207–224, 1992.
- Emmanuel Farhi and Iván Werning. Monetary Policy, Bounded Rationality, and Incomplete Markets. *American Economic Review*, (forthcoming), 2019.
- Xavier Gabaix. A Sparsity-Based Model of Bounded Rationality. *The Quarterly Journal of Economics*, 129(4):1661–1710, nov 2014.
- Simon Gächter, Benedikt Herrmann, and Christian Thöni. Trust, voluntary cooperation, and socioeconomic background: Survey and experimental evidence. *Journal of Economic Behavior and Organization*, 55(4 SPEC.ISS.):505–531, 2004.
- Jordi Galí. *Monetary Policy, Inflation and the Business Cycle*. Princeton University Press, Princeton, NJ, 2008. ISBN 978-0-691-13316-4.
- Mariana García-Schmidt and Michael Woodford. Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis. *American Economic Review*, 109(1):86–120, 2019.
- Sotiris Georganas and Rosemarie Nagel. Auctions with toeholds: An experimental study of company takeovers. *International Journal of Industrial Organization*, 29(1):34–45, 2011.
- Brit Grosskopf and Rosemarie Nagel. The two-person beauty contest. *Games and Economic Behavior*, 62(1):93–99, 2008.
- Roger Guesnerie. An exploration of the eductive justifications of the rational-expectations hypothesis. *American Economic Review*, 82(5):1254–1278, 1992.
- Roger Guesnerie. Macroeconomic and Monetary Policies from the Eductive Viewpoint. In Klaus Schmidt-Hebbel and Carl E. Walsh, editors, *Monetary Policy Under Uncertainty and Learning*, pages 171–202. Central Bank of Chile, Santiago de Chile, 2009.
- Peter Heemeijer, Cars Hommes, Joep Sonnemans, and Jan Tuinstra. Price stability and volatility in markets with positive and negative expectations feedback: An experimental investigation. *Journal of Economic Dynamics and Control*, 33:1052–1072, 2009.

- Cars Hommes. The heterogeneous expectations hypothesis: Some evidence from the lab. *Journal of Economic Dynamics and Control*, 35(1):1–24, 2011.
- Cars Hommes. *Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems*. Cambridge University Press, Cambridge, 2013.
- Luigi Iovino and Dmitriy Sergeyev. Central bank balance sheet policies without rational expectations. *Working Paper*, pages 1–63, 2018.
- John M. Keynes. *The General Theory of Employment, Interest, and Money*. Macmillan, London, 1936.
- Mel Win Khaw, Luminita Stevens, and Michael Woodford. Discrete Adjustment to a Changing Environment: Experimental Evidence. *Journal of Monetary Economics*, in press, 2017.
- Mel Win Khaw, Luminita Stevens, and Michael Woodford. Adjustment Dynamics during a Strategic Estimation Task. *Working Paper*, 2019.
- Martin Kocher, Matthias Sutter, and Werner Gueth. Experimental ‘beauty contests’ with homogeneous and heterogeneous players and with interior and boundary equilibria. *Economics Letters*, 74:219–228, 2002.
- T.C. Koopmans. On the Concept of Optimal Economic Growth. In *The Econometric Approach to Development Planning*. North-Holland, Amsterdam, 1965.
- Vivian Lei and Charles N Noussair. An Experimental Test of an Optimal Growth Model. *American Economic Review*, 92(3):411–433, jun 2002.
- Bartosz Maćkowiak and Mirko Wiederholt. Optimal sticky prices under rational inattention. *American Economic Review*, 99(3):769–803, 2009.
- Davide Marchiori, Rosemarie Nagel, and Jens Schmidt. Bounded Rationality in Interactive Decision Making: Behavioral Game Theory for Strategic Management? *Working Paper*, 2019.
- Felix Mauerberger. Monetary Policy Rules in a Non-Rational World : A Macroeconomic Experiment. *Working Paper, Columbia Economics Discussion Papers*, 2016.
- Felix Mauerberger and Rosemarie Nagel. Levels of Reasoning in Keynesian Beauty Contests: A Generative Framework. In Cars H. Hommes and Blake LeBaron, editors, *Handbook of Computational Economics 2: Volume 4*, pages 541–634. North-Holland, Amsterdam and Oxford, 2018.
- Sophocles Mavroeidis, Mikkel Plagborg-Møller, and James H. Stock. Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve. *Journal of Economic Literature*, 52(1): 124–188, 2014.
- Hervé Moulin. *Game theory for the social sciences*. New York University Press, New York, 2nd edition, 1986.
- John F. Muth. Rational expectations and the theory of price movements. *Econometrica: Journal of the Econometric Society*, 29(3):315–335, 1961.

- Rosemarie Nagel. Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5):1313–1326, 1995.
- Rosemarie Nagel, Christoph Bühren, and Björn Frank. Inspired and inspiring: Hervé Moulin and the discovery of the beauty contest game. *Mathematical Social Sciences*, pages 1–17, 2016.
- Damjan Pfajfar and Blaž Žakelj. Experimental evidence on inflation expectation formation. *Journal of Economic Dynamics and Control*, 44:147–168, jul 2014.
- F. P. Ramsey. A Mathematical Theory of Saving. *The Economic Journal*, 38(152):543, dec 1928.
- Reinhard Selten. *Preispolitik der Mehrproduktenunternehmung in der statischen Theorie*, volume 16 of *Ökonometrie und Unternehmensforschung / Econometrics and Operations Research*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1970. ISBN 978-3-642-48889-4.
- Christopher A. Sims. Implications of Rational Inattention. *Journal of Monetary Economics*, 50(3): 665–690, 2003.
- Robert M. Solow. A Contribution to the Theory of Economic Growth. *The Quarterly Journal of Economics*, 70(1):65, feb 1956.
- Dale O Stahl. Evolution of Smartn Players. *Games and Economic Behavior*, 5(4):604–617, oct 1993.
- Dale O. Stahl and Paul W. Wilson. Experimental evidence on players’ models of other players. *Journal of Economic Behavior and Organization*, 25(3):309–327, 1994.
- Dale O. Stahl and Paul W. Wilson. On Players’ Models of Other Players: Theory and Experimental Evidence. *Games and Economic Behavior*, 10(1):218–254, jul 1995.
- John B. Van Huyck, Raymond C. Battalio, and Richard O. Beil. Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure. *American Economic Review*, 80(1):234–248, 1990.
- John B. van Huyck, Raymond Battalio, and Richard O. Beil. Strategic Uncertainty, Equilibrium Selection, and Coordination Failure in Average Opinion Games. *Quarterly Journal of Economics*, 106(3):885–910, 1991.
- Carl E. Walsh. *Monetary Theory and Policy*. MIT Press, Cambridge, 3rd edition, 2010.
- Michael Woodford. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, Princeton, NJ, 2003.
- Michael Woodford. Macroeconomic Analysis Without the Rational Expectations Hypothesis. *Annual Review of Economics*, 5(1):303–346, 2013.