Levels of Reasoning in Keynesian Beauty Contests: A Generative Framework[*]

Felix Mauersberger[*], Rosemarie Nagel[†,1]

[*]*Department of Economics and Business, Universitat Pompeu Fabra, Barcelona GSE, Barcelona, Spain*

[†]*Department of Economics and Business, Universitat Pompeu Fabra, Barcelona GSE, ICREA, Barcelona, Spain*

[1]*Corresponding author: e-mail address: rosemarie.nagel@upf.edu*

CONTENTS

[*]We thank Larbi Alaoui, Jess Benhabib, Antoni Bosch, Pablo Brañas-Garza, Christoph Bühren, Antonio Cabrales, Gabriele Camera, John Duffy, Björn Frank, Willemien Kets, Anita Kopanyi-Peuker, Kiminori Matsuyama, Shabnam Mousavi, Mechthild Nagel, Antonio Penta, Michael Reiter, Ricardo Serrano-Padial, Edouard Schaal, Shyam Sunder, Michael Woodford, three anonymous referees, and the editor Cars Hommes for valuable comments. For financial support we both thank the Barcelona GSE; Felix Mauersberger acknowledges the FPI grant from the Spanish Ministry of Science and Innovation and Rosemarie Nagel the grant MINECO-ECO2014-56154-R from the Spanish Ministry of Education.

Handbook of Computational Economics, ISSN 1574-0021, https://doi.org/10.1016/bs.hescom.2018.05.002

"Because of the success of science, there is a kind of a pseudo-science. Social science is an example of a science which is not a science. They follow the forms. You gather data, you do so and so and so forth, but they don't get any laws, they haven't found out anything. They haven't got anywhere – yet. Maybe someday they will, but it's not very well developed."

(Richard Feynman, radio interview, 1981)

"[This] is the principal problem of classical economics: how is the absolutely selfish 'homo economicus' going to act under given external circumstances?"

(John von Neumann, 1932)

1 INTRODUCTION

Aiming to improve the design of cockpits, US Air Force researchers in 1950 obtained body measures from 4063 pilots, containing 140 dimensions from size to thumb length. For every dimension, the "approximate average" was defined as 0.3 times the standard deviation above or below the sample mean. For the ten most vital dimensions, between 990 (waist circumference) and 1713 (neck circumference) pilots belonged to the average defined this way. However, not a single pilot was "approximately average" with respect to all ten dimensions. Not only did the *average pilot* simply not exist, Daniels (1952) concluded that "the 'average man' is a misleading and illusory concept." As prominently argued by Kirman (1992), the "reduction of the behavior of a group of heterogeneous agents even if they are all themselves utility maximizers, is not simply an analytical convenience as often explained, but is both unjustified and leads to conclusions which are usually misleading and often wrong."

This chapter shows that laboratory experiments can be used as an effective tool to uncover the usually rich heterogeneity of behavioral responses. For our context, we define *heterogeneity* as the existence of several distinct types which have to be described using various specifications with different parameters or even different functional forms. *Homogeneity*, in contrast, means that all agents can be defined by the same specification. It is important to distinguish between heterogeneity and *randomness*: heterogeneity points to systematic differences in a specification, so that at least one structural parameter differs across the population. A situation where the population can be described by the same specification and differences only come from random draws from a well-defined distribution may more accurately be described by randomness.

We distinguish two dimensions of heterogeneity: firstly related to situations; and secondly related to behavior across games and within a game. While there is a myriad of different games and models in economics upon which experiments are often built, the benchmark solutions are typically based on an equilibrium concept, i.e. a well-defined fixed point based on maximization. Laboratory experiments, however, reveal how empirical behavior can be quite different across games: in some situations, equilibrium provides a good approximation of subjects' behavior (on average or individually) and in some it fails to do so. Within games there can be much heterogeneity among a group of agents, which is what Kirman (1992) conjectures and which is very frequently observed in the laboratory. The benchmark solution, on the other hand, might specify a homogeneous outcome for all.

In this chapter, we first review the different and disconnected areas of experimental economics within interactive decision making, each with one representative or *archetypal*, meaning "primitive" or "original" (Oxford English Dictionary, 2018) game or market and a few variations experimented on in the laboratory. We also discuss typical (heterogeneous) behavioral responses. We then offer a novel perspective by showing that the reduced forms (best response functions) of these seemingly unrelated games can be considered as special cases of a canonical game (see definition below), a generalization of the so-called Beauty Contest (BC) game. The basic BC game is characterized by a best response or optimal action as a function of others' (aggregated) actions. The word *canonical* refers to a "general rule or standard formula" (Oxford English Dictionary, 2018). Analogously, a musical canon repeats the melody in different voices. Here a canonical formulation brings together different *archetypal* games into one. In other words, all these games draw from a generative base (inspired Chomsky, 1957) that is universal.[1] The Keynesian metaphor will serve as this base, the seed, out of which emerges a multitude of models which we transform and theoretically and experimentally analyze. Examples include reduced forms (e.g. best response functions) of micro and macroeconomics, including public good games, ultimatum games, Bertrand, Cournot, some auctions, asset markets, New-Keynesian models, and general equilibrium models with sentiments/animal spirits.

The main part of this canonical form consists of the basic Beauty Contest game, which has a long history in (experimental) economics. In the original Beauty Contest game as it was first adopted by the experimental economics literature, all players "choose a number between 0 and 100 and the target or winning choice is the one closest to 2/3 times the average of all chosen numbers", where "2/3" is the main parameter of the game that can be varied. Therefore the best response function corresponds to the choice of a player equal to the target (Ledoux, 1981; Moulin, 1986); see Nagel, 1995 for the first experimental laboratory implementation).[2] The name "Beauty Contest game" was adopted by Duffy and Nagel (1997) from the Keynesian (1936) metaphor describing a contest or coordination game where newspaper readers have to pick faces which they believed to be chosen by most other readers, thus the average, the modes, or the median:

> *"[P]rofessional 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*

[1] Chompsy (1957) first used the word "generative" for a grammar as the rules guiding understandable language, which has later also been used in music theory (see e.g. Baroni et al., 1983).

[2] Nagel et al. (2017) provide a historical account of this game, showing the convergence of the macro and micro version and their applications.

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 practice the fourth, fifth and higher degrees."

(Keynes, 1936, Chapter 12.V).

Keynes speaks of several reasoning procedures, identifying unreasonable types (own taste, average opinion), and then higher order belief types of third, fourth degree and even higher. Yet, in this contest higher order beliefs collapse all in a choice corresponding to the first order belief. This is easy to see since best response to a face believed to be chosen by the average is equal to this face. Furthermore, each face can be chosen in equilibrium by all participants.

In fact, the aforementioned 2/3-of-average Beauty Contest experiment brings about behavioral belief patterns that can be described according to Keynes' idea of different levels of reasoning. In such a game, the equilibrium, where all players choose zero, is unique and can be obtained through iterated elimination of dominated strategies. All choices above $66.66 = 100 \cdot \frac{2}{3}$ are weakly dominated by 66.66. If one player chooses 66.66 and believes that others do the same, then all choices above 44.44 are eliminated, and so on, until zero will be the only choice that remains.

In contrast to this iterative process, the actual behavior observed in the experimental data can be better described by the level-k model (Nagel, 1995). This model postulates heterogeneous types and consists of one (or several) reference point(s) and (finite) iterated best replies. A reference point in interactive decision-making is typically the choice of a naive or non-strategic player, a focal point, or an intuitive choice or simple heuristics – characterized by level 0 of reasoning. In the original Beauty Contest game, the reference point is assumed to be 50, i.e. the mean of uniform random choices (that would result from insufficient reasoning). The best response to 50 is $33.33 = 50 \cdot \frac{2}{3}$, which can be referred to as level 1. A more sophisticated level-2 player may anticipate this behavior and decide to give a best response to level 1 reasoning, which would be $22.22 = 33.33 \cdot \frac{2}{3}$. A level-3 player may anticipate that behavior and best respond to that, and so on. This path can be pursued to arbitrary levels of reasoning with infinite depth in the limit, corresponding to playing the Nash equilibrium of zero.

Subjects who are not familiar with the game and the underlying theory typically select choices near or at level 0 to level 3. Therefore, also more sophisticated subjects need to respond with rather low levels of reasoning in such untrained subject pools. Thus, we provide a method attending to von Neumann's concern, how "homo economicus" should behave in external circumstances. If the Beauty Contest game is played a repeated number of rounds and subjects can observe the winning number in previous rounds, the results will be a slow convergence to equilibrium due to self-fulfilling beliefs of low levels of reasoning by other players with the reference point being the average of the preceding period. Not always is it possible to understand the reasoning process behind a choice. We therefore also discuss elicitation methods (e.g. strategy method, written comments and brain imaging studies), cognitive and population measures, to better understand heterogeneity in human reasoning in general, and in economic experiments in particular.

The level-k model and related models (Stahl and Wilson, 1994, 1995; Camerer et al., 2004; Crawford et al., 2013) have been applied to describe behavior in many different games in the laboratory and field. It has been extended and applied in behavioral microeconomics (see, e.g., Alaoui and Penta, 2017; survey by Crawford et al., 2013; Crawford, 2013), in epistemic game theory (Kets, 2012), and recently also in behavioral macroeconomics (see e.g., García-Schmidt and Woodford, 2015; Farhi and Werning, 2017).

The canonical form for these archetypal games together with a behavioral model as a framework for economic heterogeneity provides a generative structure or laws Feynman believed is missing in the social sciences. Furthermore, this framework has several advantages: first, an encompassing structure of many situations highlights why similar behavioral patterns can be identified in so many experiments on seemingly different situations across economics. In particular, if the parsimonious behavioral model does not explain behavior, one can check whether this "failure" relies on the underlying game structure or on other features, e.g., related to the subject pool. Second, this encompassing structure of situations can in turn explain why, as compared to individual decision-making, only few parsimonious behavioral game theory models have been constructed for interactive experiments.[3] Third, based on tractable relationships between identified situations and a limited number of parsimonious models, one can use observed behavior in one laboratory experiment to forecast or get an idea of subjects' behaviors in a completely different experimental setting. Fifth, the identification of a general structure may put an end to the existing repetition of experiments that conduct the same game only in slightly different contexts, and thus leaves room for the emergence of new directions in experimental economics outside this general structure, for example experiments on more complicated games.

This chapter proceeds as follows: in Section 2 we give an overview of the different areas and the newest developments in experimental economics, exemplifying heterogeneous behavior in one archetypal game within each area, and we present boundedly rational models; Section 3 provides a framework for seemingly unrelated archetypal models through a generalization of the Beauty Contest game; Section 4 discusses experimental results structured by the level-k or similar models within a systematic series of Beauty Contest and related games; Section 5 presents how different elicitation methods reveal reasoning procedures of human subjects; Section 6 concludes.

[3] We will not discuss individual decision making in this chapter. Dhami (2016) offers an excellent overview of behavioral economics, including behavioral game theory and the large area of individual decision-making. Also see Camerer (1995) for a comprehensive account on individual decision-making.

2 AN OVERVIEW OF EXPERIMENTAL ECONOMICS

This section gives a broader overview of experimental economics of interactive decision making to readers that are unfamiliar with this field. Experiments have been used as a tool in a wide range of fields in economics. While experiments with underlying research questions from different fields seem quite disconnected from each other at the first sight, we demonstrate in Section 3 that those games can be connected through a generalization of the Beauty Contest game. The reader that is familiar with experimental economics may skip this section.

In Section 2.1, we present the different areas within experimental economics in the same order as presented in the first Handbook of Experimental Economics (Kagel and Roth, 1995, vol. I). The hypotheses that are tested in the laboratory are initially motivated by economic questions specified through (game) theoretic models and equilibrium solutions based on maximization principles. Based on these hypotheses, experimenters constructed series of experiments mainly consisting of systematic changes of parameters within a few archetypal games or market structures. In some studies, the theoretical benchmarks were corroborated, while in others they were rejected. Both cases incited new work in experimental economics: when theoretical results were initially corroborated, follow-up studies were initiated testing the robustness of those results; when theoretical results were rejected, mechanisms were studied that might eventually bring about the equilibrium outcome. We present the most important games with a few variations to illustrate heterogeneity of human behavior contrasted by the (often homogeneous) theoretical solutions. In Section 2.2, we exemplify the newest developments since the late 1990s, guided by the second Handbook of Experimental Economics (Kagel and Roth, 2016, vol. II). While the first handbook covers the links between economic theory and actual behavior through laboratory experiments, in the last 20 years, experimenters have also attempted to answer questions of external validity, e.g. whether laboratory findings can be replicated in economic decision making in the real world, i.e. the field. In Section 2.3. we discuss different descriptive behavioral models of interactive decision-making that have been motivated by robust laboratory findings (see also surveys by Camerer, 2003; Crawford et al., 2013, and Dhami, 2016).

2.1 EXPERIMENTAL ECONOMICS AREAS WITH ARCHETYPAL GAMES AND MARKETS

Table 1 presents the different topics until the early 1990s. Column 1 (Table 1) states initiating questions, typically inspired by economic theory that led to the different areas of experimental economics (column 2): public goods provision, bargaining, coordination problems, auctions, financial markets, industrial organization, and individual decision making. The data collection led to the development of parsimonious behavioral models discussed in Section 2.3 (column 3).

Since subjects are typically unfamiliar with the tasks or games presented to them, most experiments include repetitions of the same game either with the same or

Table 1 Overview of experimental economics until beginning 2000s

Initiating questions	First wave: definitions of areas, series of experiments, collections of data	Second wave: development of parsimonious models, stylized facts
Inspiring theory	**Theoretically driven experiments**	**Resulting behavioral models or other developments**
Free riding vs. giving, public goods provision, punishment	Public good games	Cultural influences
Multiplicity of equilibria	Coordination games Keynesian beauty-contest games	Focal points; global game theory (noise)
Dominance solvable games	p-Beauty Contest games	Level k, learning
Fairness vs. strategic behavior Bargaining in psychology	Bargaining games (e.g. ultimatum game)	Focal points, social preferences, culture, learning models, quantal response equilibrium
Industrial organization Walrasian equilibrium	Industrial organization (Bertrand games, posted offers, etc.) Markets of buyer and seller competition	Zero intelligence
Bubbles vs. fundamental value Rational expectations	Asset markets (double auction design, call market)	Bubbles disappear with repetition of same subjects in same situation Heuristic-switching model
Revenue equivalence	Auctions	Risk aversion as explanation
Expected utility, psychology	Individual decision making	Non-expected utility models (rank-dependent U, hyperbolic discounting etc.), prospect theory, biases

changing subjects over time. This way the experimenter is able to study patterns of first-period behavior and also learning from play over time, which has been studied by a complementary theory literature (e.g. Fudenberg and Levine, 1998). Furthermore, an experimental study typically contains various treatments (experimental setups), which include different variations of some baseline treatment, also referred to as control group. This way the experimenter is able to make some causal inferences of a parameter change under ceteris paribus condition. For example, in the next section we will show the effect of punishment on behavior in public goods experiments, although the theoretical outcome is the same in a punishment and non-punishment treatment.

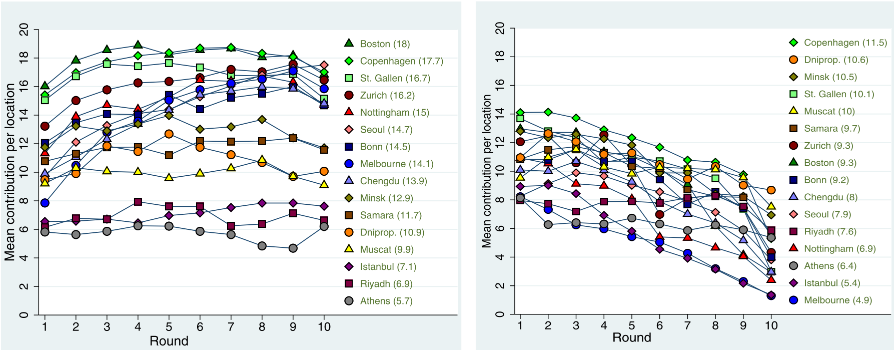

FIGURE 1

Mean contributions in a public goods game over 10 periods in two different environments. Each line corresponds to the average contribution of a particular subject pool. Numbers in parentheses indicate mean contributions (out of 20). Left: no punishment opportunity; Right: punishment opportunity (subjects can decide to punish individual subjects at the end of every round). Modified from Herrmann et al. (2008).

2.1.1 Public Good Provision

Public good games (Ledyard, 1995; Vesterlund, 2016) study situations related to the problem of free-riding. The basic game consists of N subjects within a group, who are asked to contribute all or part of their endowments. The total contribution is multiplied by a factor M ($\frac{M}{N} < 1$), redistributed equally among the members of the group and added to every individual's amount not contributed. Social optimality is attained if all players contribute their endowment to the common good. By contrast, there is an inefficient "free-riding" equilibrium with a dominant strategy, i.e. to give nothing. The main focus of the experimental results is under which conditions people contribute, especially conditionally on what others give. The initial experiments studied investigate hypotheses related to e.g., effects of group size, the multiplication parameters of the contributed amounts and information about others' contributions. A clear result is that in repeated PG games within the same group, in first rounds average giving is around 50% of the endowment but slowly decays over time.

Since the equilibrium is inefficient, experimenters and theorists need to find ways and mechanisms to deter equilibrium behavior. One effective way is punishment of free riders or low contributors. Indeed, contributions then typically do not decay but rather increase over time, despite the fact that the original equilibrium is maintained, due to costly punishment also for the punisher. This is a great example that theoretical outcomes may or may not occur in very similar setups (Fehr and Gachter, 2000).

In the last decade especially cultural heterogeneity in different societies, e.g. Russia, Switzerland, African countries etc., using students as subjects, comparing contributions, punishment, and anti-punishment (how the punished person strikes back to the cooperator) were provided. Fig. 1 shows that average contributions over time across many different countries decay without a punishment rule (left panel), while being stable or increasing over time in the presence of punishment (right panel). (Mean) contributions are very heterogeneous both across groups and within groups due to different conditional cooperation levels by individuals.[4] The authors (Herrmann et al., 2008) attribute distinct levels of contribution and effectiveness of punishment to different degrees of norm compliance in various countries. Guererk et al. (2006) construct an understudied design letting subjects decide which institution to implement: one with punishment or without. In the beginning, most groups opt for the no-punishment option, but quickly converge to the punishment rule leading to higher cooperation.

2.1.2 Bargaining

In the area of *bargaining* (Kagel and Roth, 1995; Camerer, 2003), there was a big debate on how to interpret the discrepancy between data and theoretical solutions. The *ultimatum game* (UG; Güth et al., 1982) is the archetypal game, in which a proposer has to divide a pie (amount of money provided by the experimenter), and a responder can only accept or reject the proposal. A rejection leads to both players receiving nothing and an acceptance to the distribution offered by the proposer. Any offer can be an equilibrium choice that needs to be accepted, with the lowest positive acceptable offer being the subgame-perfect equilibrium. The modal outcome in such experiments is the equal split. Small offers (below 30%) are typically rejected. In order to understand the driving forces in bargaining, many factors in the classical setup have been varied in experiments, including the number of possible offers and counteroffers over time a la Rubinstein (1982), the outside options, the number of bargainers on each side, the information about the pie size etc. The simplest version of a bargaining game in this context is the "dictator game" (without responder behavior).

Camerer and Fehr (2006) attribute the heterogeneity in behavior to a dichotomy of types in the population that is inspired by theoretical biology (Gintis, 2003): self-regarding individuals, who act without applying a particular notion of fairness, and strong reciprocators, who are characterized by giving others altruistic rewards for cooperation and norm-conformity, and applying altruistic punishment in case of norm-violations.

Fig. 2 shows average results in ultimatum games with subjects from different cultures and small scale societies by Henrich et al. (2001). The authors argue that different market and trading rules among the members can identify the different patterns of behavior. For instance, the Lamelara, being predominantly a whale-hunter culture, rely on cooperation for daily fishing, which might explain higher offers in ultimatum games. On the other hand, small scale societies who experience less coordination among group members like Machigueng offer much less when being proposers. In Section 2.3 we discuss the fairness models that rationalize such behavior with social preferences.

[4] See Arifovic and Duffy (2018) for further details.

Group	Country	Mean offer[a]	Modes[b]	Rejection rate[c]	Low-offer rejection rate[d]
Machiguenga	Peru	0.26	0.15/0.25 (72)	0.048 (1/21)	0.10 (1/10)
Hadza (big camp)	Tanzania	0.40	0.50 (28)	0.19 (5/26)	0.80 (4/5)
Hadza (small camp)	Tanzania	0.27 (38)	0.20 (8/29)	0.28 (5/16)	0.31
Tsimané	Bolivia	0.37	0.5/0.3/0.25 (65)	0.00 (0/70)	0.00 (0/5)
Quichua	Ecuador	0.27	0.25 (47)	0.15 (2/13)	0.50 (1/2)
Torguud	Mongolia	0.35	0.25 (30)	0.05 (1/20)	0.00 (0/1)
Khazax	Mongolia	0.36	0.25		
Mapuche	Chile	0.34	0.50/0.33 (46)	0.067 (2/30)	0.2 (2/10)
Au	PNG	0.43	0.3 (33)	0.27 (8/30)	1.00 (1/1)
Gnau	PNG	0.38	0.4 (32)	0.4 (10/25)	0.50 (3/6)
Sangu farmers	Tanzania	0.41	0.50 (35)	0.25 (5/20)	1.00 (1/1)
Sangu herders	Tanzania	0.42	0.50 (40)	0.05 (1/20)	1.00 (1/1)
Unresettled villagers	Zimbabwe	0.41	0.50 (56)	0.1 (3/31)	0.33 (2/5)
Resettled villagers	Zimbabwe	0.45	0.50 (70)	0.07 (12/86)	0.57 (4/7)
Achuar	Ecuador	0.42	0.50 (36)	0.00 (0/16)	0.00 (0/1)
Orma	Kenya	0.44	0.50 (54)	0.04 (2/56)	0.00 (0/0)
Aché	Paraguay	0.51	0.50/0.40 (75)	0.00 (0/51)	0.00 (0/8)
Lamelara[e]	Indonesia	0.58	0.50 (63)	0.00 (3/8)	0.00 (4/20)

Note: PNG = Papua New Guinea.
[a] This column shows the mean offer (as a proportion) in the ultimatum game for each society.
[b] This column shows the modal offer(s), with the percentage of subjects who make modal offers (in parentheses).
[c] The rejection rate (as a proportion), with the actual numbers given in parentheses.
[d] The rejection rate for offers of 20 percent or less, with the actual numbers given in parentheses.
[e] Includes experimenter-generated low offers.

FIGURE 2

Ultimatum game in small scale societies. Source: Henrich et al. (2001).

In Section 4 we will show disaggregated behavior in UGs both in the first period and after learning has occurred; and in Section 5, we show results for incomplete information UGs.

2.1.3 Coordination Games

Coordination games (see also Ochs, 1995; Camerer, 2003) will be the main focus of attention in this chapter in the following sections with the Keynesian Beauty contest as its corner stone. Behavior in such games is particularly prone to bringing about heterogeneity due to the multiplicity of equilibria, often being Pareto-ranked (some equilibria give all members higher payoffs than other equilibria which, however, might provide secure payoffs or are risk-dominant). Thus, even rational players cannot coordinate a priori to a single prediction by introspection. Therefore, experimentation can be useful and necessary to periodically study equilibrium selection. A simple example is whether we meet at the Empire State Building or at Grand Central Station in New York. Two strangers might resolve the coordination problem differently than two New Yorkers. The focus of the literature on coordination games has offered potential solutions. In general groups of 2–3, maybe also 4 individuals coordinate rather well, while groups larger than 4 typically converge to inefficient equilibria or fail to coordinate altogether. Experimental researchers introduce different methods that subjects also in large groups achieve coordination, e.g. by means of communication. The theoretical literature, e.g., shows that payoff perturbations can lead to a unique equilibrium (initiated by Carlson and Van Damme, 1993 on global games), which we will discuss in Sections 3 and 4.

2.1.4 Industrial Organization

Industrial organization was one of the early areas in experimental economics, starting with Sauermann and Selten (1959) on complicated oligopoly games with demand inertia that led to the theoretical paper of subgame perfection Selten (1965).[5] Experimental industrial organization is surveyed e.g. by Brandts and Potters (2017), Holt (1995), and Davis and Holt (1993). The main topic of this area studies the question of price formation and production in monopolies, oligopolies, and Walrasian markets through different market institutions. Plott and Smith (1978) study the emergence of prices and gains from trade in a private goods market. In the experiments, discrete supply functions are constructed by inducing privately known costs to sellers, who all possess one unit of a good, and demand functions by giving privately known reservation values to buyers who all want to buy one unit of the good. Thus, there is exogenous heterogeneity of types (seller vs. buyer) and sellers or buyers, respectively, can have different values. In the complete information efficient equilibrium the price is such that supply equals demand.

One possible trading mechanism is the double auction, introduced by Smith (1962), in which buyers' bids and sellers' asks together with the resulting contract prices are displayed publicly on a blackboard or on participants' screens. Once an announced bid is higher than an announced ask, trade occurs. The resulting prices converge fast to equilibrium which might be the best example of theory coinciding with observed behavior. The equilibrium in this context is the Pareto-optimal Walrasian equilibrium equating demand and supply and high trading efficiency.

[5] See also a discussion by Abbink and Brandts (2010), Nagel et al. (2016).

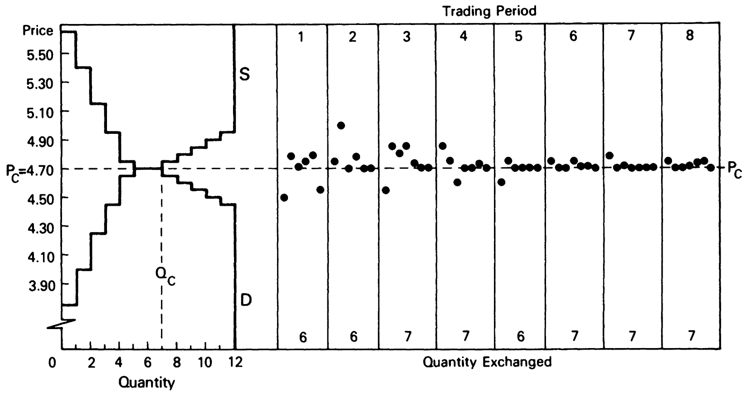

FIGURE 3

Contract prices (dots) in a double auction and the Walrasian equilibrium (datched lines) over time, with supply and demand schedules (left side); from Ketcham et al. (1984).

Fig. 3 shows demand and supply functions of a market (left side of graph) and the resulting contract prices within different time intervals. Only in the first trading periods most contract prices are off equilibrium. However, from period 4 onwards, within a period only initial contract prices are below (above) the equilibrium, and then players coordinate on the equilibrium price. Notably, not all theoretical requirements need to be fulfilled: e.g. a handful of agents on each side suffices instead of an infinite number of agents; demand and supply functions do not need to be commonly known among acting agents, but instead each agent only knows her own cost or willingness to pay.

Gode and Sunder (1993) provide a simple zero intelligence model in which computer programs randomly choose bids and asks (thus, producing high heterogeneity of behavior), constrained by own cost, reservation values and the trading rules. They find that markets populated by automaton traders following such simple rules will converge quickly to the equilibrium. However, the distribution of the surplus depends on the intelligence of traders, giving more to those who are more sophisticated. Thus, humans can easily exploit such mechanical traders. These results are fairly robust to parameter changes. Neither shifting supply or demand curves, reducing the number of sellers to the monopoly case, providing insider information, or incomplete information about the true state of the world offsets the Walrasian equilibrium behavior in most repetitions.

2.1.5 Asset Markets

Sunder (1995) discusses *asset markets* as stylized situations of financial markets, studying information asymmetries between traders. Sunder (2007) highlights that a key development in experimental finance nowadays are experiments to understand the causes and sources of financial bubbles, while before the 2000s, experimental finance dealt with the acquisition and aggregation of information in financial markets.

In the simplest case (Smith et al., 1988) all subjects are identical, endowed with the same number of shares, money that pays no interest, and information about fundamental values (FV), dividend payments of shares per period, and a finite horizon after which the shares are worthless. Trading is done via the double auction, explained in the previous section, or a call market. In the later orders are cleared by a single price, equating demand and supply constructed through the bid-quantity and ask-quantity tuples of participants. Everyone can choose to be a buyer or seller of shares. In reality, it is difficult to separate the fundamental value of an asset apart from the bubble component of realized prices.

Fig. 4 shows a typical outcome of such an asset market played over 15 periods with an occurring bubble. The known fundamental value decreases over time, being zero after 15 periods (see horizontal lines, rational expectation prices). Only the asset pays a dividend while money holdings do not. The solid increasing and decreasing lines represent the realized contract prices in each period and the other dots are subjects' asks and bids, which are quite dispersed although all subjects are ex-ante identical.

Haruvy et al. (2014) conduct bubble experiment and classify players into 3 types based on De Long et al. (1990): fundamental (who are guided by fundamental values), speculative (who exploit momentum traders), and momentum traders (who follow the trend of the contract prices) to explain the presence of bubbles.

If the same group of subjects plays again in the same setup (a multiple period market as described above), once or twice, maintaining the same parameters, the bubble disappears and behavior becomes rather homogeneous with respect to bids and asks. Systematically changing fundamental values (decreasing, increasing, or constant over time), introducing interest rates for money holding results in bubbles. Yet, increasing or constant FVs result in prices closer to the FV than decreasing FV, suggesting that the later produces more confusion in subjects than the former (Breaban and Noussair, 2015). In this paper the authors also test and analyze the influence of individual characteristics (such as risk aversion, loss aversion, cognitive measures) on pricing behavior.

Based on the large amount of experiments with such financial markets, it is easy to predict whether or not a bubble occurs – e.g. the degree of strategic complementarities in the underlying system and the characteristics of subjects are decisive factors. However, as in the real world, one cannot predict the magnitude or time period of the bursting of the bubble. Yet, as mentioned above, one can predict under which conditions bubbles should be smaller or bigger (on average). (See also the discussion on asset markets by Arifovic and Duffy, 2018.[6])

2.1.6 Auctions

Auction experiments (Kagel, 1995) represent the best interaction between theory, laboratory, and reality. Think of a restaurant owner, who buys her fish, seeing the fish

[6]In 2010 researchers working in this area created the *Society of Experimental Finance*.

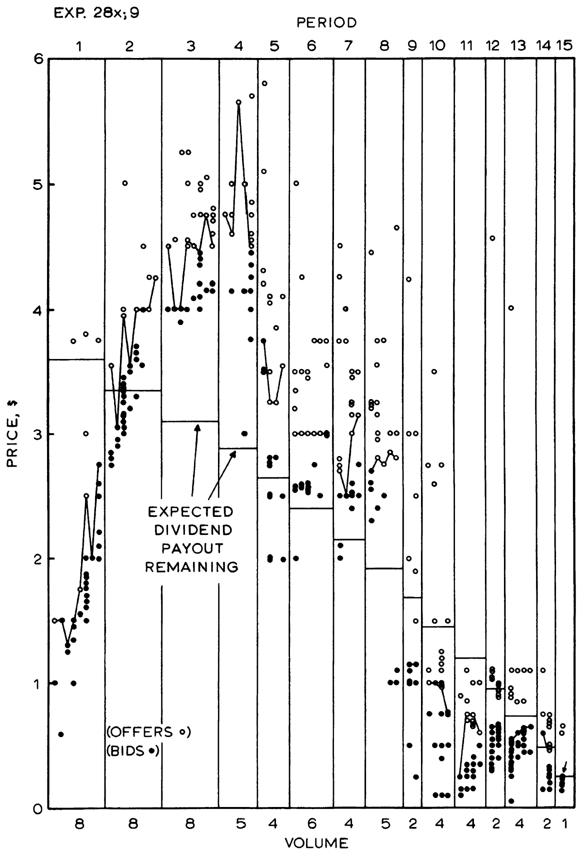

FIGURE 4

Chart of an asset market with all bids and offers and the resulting contract prices (joined by line segments) in sequence for experiment session 28x showing the price dynamics both within and across trading periods in one market. The horizontal lines present the fundamental values for each period. Source: Smith et al. (1988).

coming into the auction hall, and competing with the daily returning other buyers. She clearly knows the auction rules and her willingness-to-pay. As stated already above, understanding or testing behavior related to mathematical models with lab experiments has some advantages over naturally occurring data in the field. Here, certain parameters can be induced by the experimenter as drawing the privately known willingness to pay of each participant from a known, given distribution, in order to calculate a private-value equilibrium. This allows to precisely know deviations of behavior from rational solutions. The first main focus of laboratory experiments in this area consisted of comparing realized revenues and behavior between different auction rules in private and common value single good auctions. Theoretically, average revenues are the same in certain classes of auctions. Kagel and Levin (2002) review

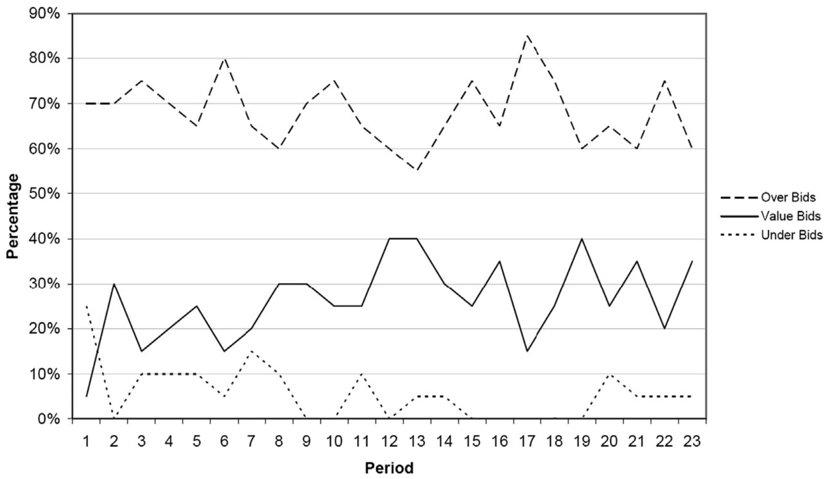

FIGURE 5

Kagel and Levin (1993): Second price auction data as reanalyzed by Garrat and Wooders (2010); bid $< \pm\$0.05$ of value $=$ value; $\frac{2}{3}$ of bids are above private value, although this is strictly dominated.

the influence of the requirements in the newly arising multi-unit auctions starting with the Federal Communications Commission (FCC) auction in the US since 1994. The FCC realized that spectrum auctions more effectively assign licenses than either comparative hearings or lotteries. Here, theorists and experimenters greatly interacted to solve challenging questions such as which auction types to implement or how to deal with bundling of objects.

Let us exemplify heterogeneous behavior in a second-price private value auction with a dominant strategy equilibrium (bidding own private value). All players receive a private value, i.i.d. drawn from a known distribution. The highest bidder will receive the object, paying the second highest bid. The same theoretical results hold for the English auction in which a player quits when an ascending clock reaches her private value, and the last remaining player receives the object for a price determined by the value at which the second last bidder dropped out.

Fig. 5 presents behavior over time for a second price auction categorized by three types: overbidding, i.e. bidding above one's private value (approximately 70%); value bidding (20% in the beginning and then 30%); and underbidding, i.e. bidding below one's value (10% in the beginning and then 0%). Players do not converge to the simple dominant strategy equilibrium over time. This is in contrast to behavior in English auctions, where subjects drop out once their value is reached by the ascending clock. The reasoning procedure for finding the equilibrium bid is much harder for the second price than for the English auction. In the later potential losses are salient, when staying in, once the clock surpasses the private value. However, in the second price auction, a player seems to hope to receive a lower price than her own value and at the same time to augment the chance of winning by an increase of the bid. Since typically subjects do not incur losses, they cannot learn to avoid such wrong reasoning. Similarly, common value auctions produce overbidding and the winner typically pays a much higher price than her private value (called winner's curse). This is due to the difficulty in understanding the theoretical implications of correlated private signals of the common value. Furthermore, players do not learn to avoid such errors neither in the field nor in the laboratory. In the laboratory, each period a new common value together with correlated private values is drawn. An environment like this one being subject to unremitting changes does not allow to draw clearcut implications about the optimal bidding behavior.

2.2 NEW QUESTIONS AND AREAS SINCE THE LATE 90s

Many critiques about experimental economics have repeatedly been expressed, typically from outside this field. The three main concerns were the choice of subject pools (mainly undergraduate students at universities), the question of external validity of laboratory results, and low payoffs. For this purpose, the robustness of laboratory experiments with undergraduates has been tested using different subject pools such as chief executive officers, traders from the stock market, etc. (See example of the ultimatum game above.) Recently, Kagel and Roth (2016) issued the second volume of the Handbook of Experimental Economics containing the newest developments with these and other kinds of questions which we briefly summarize in the remainder of this section.

Field Experiments

All chapters of Kagel and Roth (2016) discuss field experiments, probably the most prevalent experimental methodology besides laboratory experiments (see also survey by Harrison and List, 2004; Banerjee and Duflo, 2018). A field experiment contains a targeted intervention by a researcher in the real world for the purpose of collecting appropriate data to answer a research question. In a field experiment, the researcher does not typically create an artificial setting that she controls but exerts control over some environment or some aspect of subjects' real life. This is usually done by randomly dividing the subject pool into a "treatment" group, being subjected to an intervention or policy, and a "control group", being a comparable group not undergoing that intervention.[7] The challenges of field experiments are usually categorized into "internal validity" considerations, referring to concerns how to consistently estimate the treatment effect, and "external validity" considerations, referring to questions to what extent one can generalize the results of the field study (see Banerjee and Duflo, 2009 and Duflo, 2006 for more detailed discussions). The Internet has become experimenters' playground to set up such natural experiments. For example, Hossain and Morgan (2006) introduce different shipping and handling charges to examine

[7]Consider in contrast to that a natural experiment, where the researcher uses some event that happens to randomly divided group of interest into "treatment" and "control."

Table 2 New questions and new areas of experimental economics

New questions after the collection of data	Third wave: leaving the economic laboratory and other new areas
External validity Policy implications Incentives (high, low) Subject pools: experts vs. students Influences of socio-economic variables	Field experiments (RCT) e.g. auctions in the fields, FCC Auctions (multi-unit auctions and bundles of auctions) Incentives field vs. lab Development experiments ⇒ policy implications Experiments with different populations (children, small scale societies, professionals e.g. football players etc.) Cross-cultural experiments
Risk in interactive decision making	Price Lists (a series of ordered binary lottery choices) combined with behavior in games
Unraveling in markets	Market design Matching in the field
Male vs. female; happiness, ethics	Gender experiments, personal characteristics, ethics (lying experiments) and happiness surveys, cognitive IQ, strategic IQ
Expectation formation, forecasting, discounting, policy implications, central bank communication	Macroeconomics, finance, political economics experiments ⇒ policy implications
How can biological data inform decision making?	Neuroeconomics (especially individual decision making) lesion patients, animals vs. humans; new technologies from the natural sciences: e.g., fMRI, eye-trackers, skin conductance

shopping behavior. Entirely different field experiments introduce (network) structures using Facebook and other social media or online experiments via Mechanical Turk. (For methodological questions about such experiments, see Chen and Konstan, 2018.) Yet, doing the abstract games we discussed above with subject pools different from students also constitute field experiments. As a summary, in the laboratory, many features are controllable by the experimenters, i.e., the induced values of an abstract object in an auction, which in the field is typically known to the subject himself for the real object.

Table 2, column 1, states the main new questions and challenges: *macroeconomic experiments* (Duffy, 2016; see also the chapter by Arifovic and Duffy, 2018); *political economy experiments* with a focus on committee bargaining and voting issues (Palfrey, 2016); Roth (2016) discusses *market design* related, e.g. to school choices, labor markets, kidney exchange, etc. While unraveling in public good games can be offset through punishment, markets such as labor markets or school choices also suffer from inefficiencies, if the institutions are not well designed. Roth compares different matching mechanisms. The most prominent one is the Gale–Shapley (1962) algorithm, with which, for example, unraveling of early matching within a market is largely offset. Frechette (2016), summarizing studies with different subject pools, such as experts, children[8] and even animals, concludes that results are mostly similar across different kinds of subjects. Kagel and Levin (2016) discuss newest developments in *auction* designs and data, especially raised through questions of Federal Communications Commission (FCC) auction designs introduced in the mid-90s, involving multi-unit auctions. Erev and Haruvy (2016) discuss *learning* in simple decision tasks.

The new field of *neuroeconomics* (Camerer et al., 2016) is a first start to bridge the gap between experimental economics and the natural science, linking biological data like brain activity or eye-movements to the behavioral data within existing or new economic experiments. We will discuss this field with a few examples at the end of Section 5.

Gender Experiments

Gender issues have constituted a large topic in empirical economic research, stressing in particular differences in outcomes as wages between men and women based on discrimination. Niederle (2016) discusses heterogeneity of behavior driven by gender and race, taking a different approach. The main result of this experimental topic seems to be that women dislike competition, first documented in Gneezy et al. (2003). They show that this dislike is a main difference between men's and women's entry rates in tournament games (only the winners receive a payment). The reason is related to women's underconfidence of own performance, which contrasts with men's overconfidence. The tasks in such experiments are, for example, about solving mazes or simple calculations in which men and women roughly earn the same with performance-based payments. Fig. 6 shows the percentage of entry rates into standard tournaments (ST) based on the probability of winning the tournament. In general men's entries are much higher than women's. Affirmative action (AA, reserving one of the two winner positions for the highest performing women) considerably helps improve entry by women and decrease men's entry rates.

2.3 BOUNDEDLY RATIONAL MODELS

In the beginning of the 90s, behavioral economics emerged with several new modeling tools of structuring actual behavior in a consistent way (see Table 1, column 3) in order to understand the relationship between actual behavior and theory across the different areas mentioned in the previous section. The special issue in *Games and Economic Behavior* (1995) documents the shortcomings of fully rational concepts

[8]See also Harbaugh and Krause (2000) who started to conduct research on experimental games with children, showing, e.g., when theory of mind develops in children in sharing tasks.

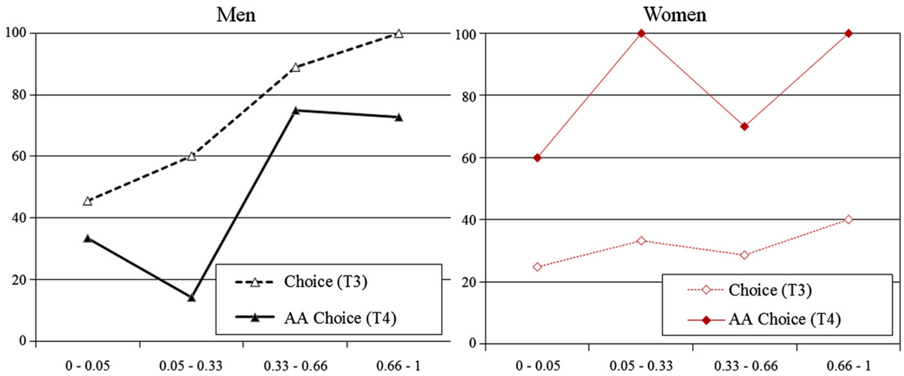

FIGURE 6

Percentage of participants entering the tournaments as a function of probability of winning the specific tournament (T3 is the standard tournament (ST) and AA T4 is the tournament with affirmative action). Modified from Niederle et al. (2013).

such as Nash equilibrium (Nash, 1951) in describing behavior observed in experiments and presents the first important alternatives:

Learning Models

Given a large body of data in the area of bargaining models and market or auction, Roth and Erev (1995) develop reinforcement learning, a learning mechanism originating from psychology, into economic games.[9] This model requires low rationality, without the need to know the precise environment of the game. Subjects update their propensity to play a given strategy through the payoff feedback of their actual past actions. This theory is quite different from the Bayesian learning literature dominating the field of theoretical learning in the game-theoretic and econometric literature. However, both reinforcement learning and Bayesian learning (including fictitious play, which has a Bayesian foundation) have been combined in a more general model, the experience-weighted attraction (EWA) model, by Camerer and Ho (1999). Subjects not only update propensities of chosen but also they keep track of hypothetical payoffs from actions that are not played, which requires more knowledge of the environment than the basic reinforcement model. Camerer (2003) and Erev and Haruvy (2016) provide an overview of such models (see also chapter by Arifovic and Duffy, 2018).

According to the basic forms of reinforcement learning and EWA, players follow the same learning rule regarding the way they make decisions. Stochasticity in those

[9]Reinforcement learning models are also found in the machine learning literature. Other examples of algorithms found in the machine learning literature that have been used in economics are hill climbing (Nagel and Vriend, 1999) and Thompson sampling (Mauersberger, 2018).

models is a primitive form of accounting for variation in decision-making over time or in the cross-section.

Quantal Response Equilibrium Model (QRE)

At the same time the so-called quantal response equilibrium model was developed by McKelvey and Palfrey (1995), orthogonally to the learning literature. Agents make mistakes (as their hands may tremble; Selten, 1975) in choosing a best response. By doing so, they form equilibrium beliefs about their opponents that are true in expectation (see survey by Goeree et al., 2014).

Like in reinforcement learning and EWA, players are homogeneous in the basic form of QRE in the way they make decisions. However, quantal response equilibrium model is a stochastic model, allowing decisions to differ over several dimensions such as over repetitions of the game, and the cross-section. Heterogeneity in observed behavior can emerge through random draws. For example, if players start in the same way, stochasticity can push them into different directions and thus also create different payoffs. Rogers et al. (2009) and Breitmoser (2012) expand quantal response equilibrium to include heterogeneity by introducing individual-specific λ-parameters into a logistic equilibrium.

Level-k

The level-k model is an alternative boundedly rational reasoning model (see survey by Crawford et al., 2013), which contains heterogeneity at its core based on the idea that subjects have different degrees of sophistication as mentioned before.[10] This model was developed by Nagel (1995) within the beauty contest experiment and is discussed together with related models in more detail in Section 4.1.2. A requirement or assumption of this model is that the decision-maker knows the strategic environment of the game. This means that it is not obvious that level-k prevails if the decision-maker is not given exact knowledge about the game's parameters. In fact, a large literature (see e.g. Gigerenzer et al., 1999; Hommes, 2013) documents that individuals use heuristics in the absence of precise information about their environments. Heterogeneity can still be observed, because individuals may adopt quite different heuristics. While level-k can be considered a particular heuristic itself, one specific heuristic element contained in level-k thinking is level 0, often corresponding to focal points or reference points. (See e.g. Nagel, 1995; Fehr et al., 2017.)

Social Preference Models

Social preference models cover a fair amount of behavioral models to explain deviations from rational choice models (see surveys by Camerer, 2003 and Cooper and

[10] An earlier approach based on stepwise reasoning was eductive learning (Guesnerie, 1992), which investigates whether agents learn an equilibrium by a process of introspection. In contrast to that, level-k postulates that agents use finite iterated steps of reasoning.

Kagel, 2016). The main feature is the expansion of the utility function by specifically considering the preferences or intentions of other individuals. The difference to boundedly rational models is that they require no mistakes or "trembles" such as in the QRE model or in the reinforcement model and are based on the assumption of consistent beliefs as QRE. Therefore, social preference models cannot explain non-maximizing rejection rates as they should not happen in equilibrium. Examples for social preferences include reciprocity, describing the intrinsic desire of an individual to treat other individuals according to how they treat this individual (Rabin, 1993; Falk and Fischbacher, 2006; Dufwenberg and Kirchsteiger, 2004); inequity aversion, describing the aversion against inequalities both in favor or against a particular individual (Fehr and Schmidt, 1999; Bolton and Ockenfels, 2000); and quasi maximin, describing the notion that a decision-maker may care about society's average payoff and the minimum payoff apart from her own payoff (Charness and Rabin, 2002). The literature shows that individuals are heterogeneous regarding their development of such social preferences. Heterogeneity comes about through different parameter constellations in the same or across populations, e.g. the intensity how much you care about your own payoff versus the payoff of the other player(s).

Crawford et al. (2013) discuss the boundedly rational models level-k and quantal response equilibria and others. Crawford (2013) distinguishes between two types of models deviating from full rationality: the first category is "boundedly rational" models, including frameworks that relax the assumption of individual optimization; the second one is "optimization-based" models, which presume some notion of optimality but relax any other assumptions than optimization. While the first category includes models like reinforcement learning, the second category includes models of reference-dependent preferences (Kahneman and Tversky, 1979; Tversky and Kahneman, 1991; Koszegi and Rabin, 2006), models of limited information about the causal structure of the environment to allow for heuristics and biases (Tversky and Kahneman, 1974; Rabin, 2002), and behavioral game theory models of learning such as adaptive models (Woodford, 1990; Milgrom and Roberts, 1990, 1991; Selten, 1991; Crawford, 1995; Fudenberg and Levine, 1998; Camerer and Ho, 1999; Camerer et al., 2002), level-k (Crawford et al., 2013), and "cognitive hierarchy" models (Camerer et al., 2004). Crawford argues that models of the second category (relaxing other assumptions than optimization) are superior to purely boundedly rational models in explaining common phenomena in microeconomics such as systematic overbidding in auctions, bubbles and crashes and finance, the winner's curse and informational naiveté, the latter usually referring to situations where not all available information is used.

3 THE KEYNESIAN BEAUTY CONTEST: A GENERATIVE FRAMEWORK FOR ARCHETYPAL GAMES IN ECONOMICS

In this section we develop a new concept encompassing seemingly different models into one framework. While this section is far from being complete, we hope to inspire a fruitful discussion of structuring models with the aim to make better behavioral predictions in the laboratory and in the field supported by empirical and theoretical validations.

One strength of experimental economics is the systematic control over the parameters of a game or a class of games to test compliance with theoretical predictions, causality issues, and to study behavioral regularities or behavior over time. We choose an approach that helps explain the emergence of heterogeneity by the game structure, i.e. aspects of the environment that are under the control of the experimentalist. Many of the features of the games we bring together in this section have been discussed as overarching themes of common structures by other researchers (as e.g. strategic substitutes and complements and discrete vs. continuous strategies). However, the new idea here is to show that seemingly unrelated situations, some studied in Section 2, can largely be encompassed as special cases of a general function, which is also generative (Chomsky, 1957).

We combine the original *Beauty Contest game*, which is a best response function of aggregated opponent choices, with two additive terms: 1) a constant (representing e.g. a fundamental value of a share or a pre-announced choice of a player) and 2) an idiosyncratic random term (see also Benhabib et al., 2015).[11] This linear function becomes a canonical form for including many seemingly different archetypal models or their abstract simplifications as special cases: e.g. the general equilibrium model with sentiments/animal spirits, New-Keynesian model, asset markets, Cobweb, Cournot reaction functions, ultimatum games, some auctions, and also discrete choice games as stag hunt games, market entry games.

We hope such a common functional form across different models together with level-k will encourage new theoretical, behavioral, and experimental research, to make better predictions that are useful for designing better institutions in the real world. Once we understand the more general structure of those strategic interactions we can more easily understand the driving forces of human behavior and the heterogeneity that may emerge even if the theoretical solutions suggest homogeneous outcomes.

We will show in Section 4 how parameter changes that do not affect the equilibrium outcome may change actual out-of-equilibrium behavior as observed in the laboratory. Conversely, equilibrium differences might not induce behavioral changes due to subjects' cognitive limitations.

3.1 THE BEAUTY CONTEST AS A CANONICAL FORMULATION

Definition 1. Suppose there are N individuals and, at any time $t = 1, ..., T$, every individual i chooses (simultaneously to the N-1 other individuals) an action $y_t^i \in \mathbb{Y}^m$, where $\mathbb{Y}^m$ is an m-dimensional space. We refer to a game as a *Beauty Contest game*

[11] Our endeavor resembles Bergemann and Morris (2013). However, while they start with a similar general equation as ours, they are interested in different information structures within their formulation and resulting equilibria.

Table 3 Classification of archetypal games

<table>
<tr><td colspan="3">Optimal choice (best response) for player i:
$y_t^i = \hat{E}^i\{c_t + b \cdot f(y_t^1, ..., y_t^N) + d \cdot f(y_{t+1}^1, ..., y_{t+1}^N) + \epsilon_t^i\}$</td></tr>
<tr><td></td><td>Anti-coordination games
Strategic substitutes
Negative feedback
f decreasing function</td><td>Coordination games
Strategic complements
Positive feedback
f increasing function</td></tr>
<tr><td rowspan="4">1. Continuous choice variable:
from interval (e.g. [0,100]);
or any number</td><td colspan="2">either substitutes or complements:
Basic case: Beauty Contest game ($b < 0$ or $b > 0$)[a]
General case:
General equilibrium model with sentiments/animal spirits ($b < 0$ or $b > 0$)</td></tr>
<tr><td colspan="2">strategic heterogeneity:
for some players $b < 0$ and for some $b > 0$</td></tr>
<tr><td>Cournot game

Ultimatum game
Cobweb model</td><td>Bertrand game
Minimum/median effort game
some auctions
BC with (un)known fundamental
Asset market (learning to forecast)
New-Keynesian model (learning to forecast)[b]</td></tr>
<tr><td colspan="2">neither substitutes nor complements: public goods, two-person BC with tournament payoff, harmony game (dominant strategy)</td></tr>
<tr><td rowspan="3">2. Discrete choice variable:
from interval (e.g. [1,2,3...,7];
or [A,B])</td><td>Lowest unique bid auction (LUPI)
Entry (chicken) game

Global games (congestion)
Negative assortative matching</td><td>Keynesian Beauty Contest
Stag hunt game
Battle of Sexes
Global games (Attack or not)
Positive assortative matching</td></tr>
<tr><td colspan="2">strategic heterogeneity: matching pennies (hide and seek), fashion cycles</td></tr>
<tr><td colspan="2">neither substitutes nor complements: prisoner's dilemma, harmony game (dominant strategy)</td></tr>
</table>

y_t^i: choice of individual i in period t; $\hat{E}_t^i$: subjective (possibly non-rational) belief of individual i in period t; $f(.)$: aggregation function (linear for the continuous case; we allow step functions for the discrete strategy space).

[a]To avoid confusion with other variables such as price, we call this parameter b in our chapter, instead of p as in the experimental literature.

[b]Degree of positive feedback depends on interest rate rule.

if individual i's best response/optimal action is:

$$y_t^i = \hat{E}_t^i\{c_t + b \cdot f(y_t^1, ..., y_t^N) + d \cdot f(y_{t+1}^1, ..., y_{t+1}^N) + \epsilon_t^i\} \tag{1}$$

where $b, d \in \mathbb{R}^{m \times m}$. $c_t \in \mathbb{R}^m$ is a common, deterministic, possibly time-varying process, $\hat{E}_t^i$ denotes the subjective belief of individual i held at time t and $\epsilon_t^i \in \mathbb{R}^m$ is an idiosyncratic exogenous, stochastic process.

$f^{(i)}(.) : \bigtimes_{j=1}^N \mathbb{Y}^m \rightarrow \mathbb{Y}^m$ produces a vector $y^{agg,i} \in \mathbb{Y}^m$ whose q-th element $(q = 1, ..., m)$ contains either the average, minimum, maximum, median, mode, action least chosen, an action chosen by at least $h \leq N$ agents or the sum (provided they exist) of the q-th elements of all individuals' $(j = 1, ..., N)$ present action vectors y_t^j or future action vectors y_{t+1}^j.[12] If $f(.)$ uses the mode, the action least chosen or an action chosen by at least $h \leq N$ agents and that statistic is not unique, the one that gives individual i the higher payoff is chosen for $y^{agg,i}$.

In many cases, $c = c_t, \forall t$ reduces to a constant. We limit ourselves to the average, minimum, maximum, median, mode, action least chosen or the sum as aggregation rules, since they are of great importance for economic games and markets.[13] While optimality conditions like Eq. (1) are normally solved by concepts like rational expectations or Nash equilibrium, other less standard solutions concepts are possible of which we discuss one in detail later in Section 4: level k. From Section 2, we will reformulate the public good game, the ultimatum game, several coordination games, an asset pricing market, an auction, and several other games not discussed in detail before.

We present examples from the economic theory literature and the corresponding experimental literature. Note that (1) can describe the optimal action (or best response) for games and setups that differ in several dimensions. Table 3 only contains four of the most important dimensions:

1. **Continuous vs discrete choices:** When strategies are chosen from a finite, discrete set, the typical examples are normal form games with 2 or more strategies; when there are 2 players the experimental literature has essentially defined only five important symmetric games: the prisoners dilemma, chicken game, stag-hunt game, matching pennies, and the harmony game.[14] Contrary to that, strategies can be chosen from a continuous set, i.e. from real numbers, which would be the case in an ultimatum game, an investment or consumption game where you decide the amount to invest or to consume, and in the standard p-Beauty Contest

[12]The sign after c need not be positive, as a negative sign can be introduced by b or d.

[13]This kind of aggregation excludes, for example, some asymmetric three player games, in which a player cannot be represented by a constant or the two opponents with a simple aggregation rule. Furthermore, we confine ourself to a linear function as all the games we discuss have this form. Arifovic and Duffy (2018) discuss the non-linear case by Fehr and Tyran (2001, 2007, 2008).

[14]These five symmetric games essentially entail the important characteristics of conflict and cooperation as dominant strategy, multiplicity of equilibria, (in)efficiency of equilibria. There is a long tradition to categorize the important 2x2 games in psychology and economics (see e.g. Rapoport et al., 1976).

game. A Bertrand or public good game is just the continuous version of a PD game. We also will show the effects on behavior when choices can be from the entire real numbers or contrary when the choice set is bounded.

2. **Strategic complements (coordination) vs strategic substitutes (anticoordination):** Loosely speaking, a game of strategic complements is one where an individual has to match others' actions and can thus also be called coordination game. Conversely, a game of *strategic substitutes* is one where an individual i has to act in an opposite way towards others so that the game can alternatively be called anticoordination game (see e.g. Vives, 2010). In the context of the Beauty Contest game, choices are complements (substitutes) if $f(.)$ is an increasing (decreasing) function of another individual's choice y^{-i}. Speaking of "increasing" and "decreasing" highlights an important assumption of an ordered strategy space that is often imposed in case of strategic substitutes and complements. The case of increasing (decreasing) functions is often referred to as positive (negative) feedback, as stressed in Heemeijer et al. (2009).
This terminology can serve as a means of classification for economic environments in both microeconomics and macroeconomics. Bertrand games are commonly characterized by strategic complementarities, because if firms decrease prices, other firms have an incentive to curb prices as well. The Cournot model is, on the other hand, a good example of strategic substitutability, as the higher the quantity a particular firm produces, the lower the quantity another firm produces. In macroeconomics, strategic complementarities play a role in the presence of search frictions (Diamond, 1982), information frictions (Bryant, 1983) and increasing returns (Weitzman, 1982). See Cooper (1999) for an early book for the role of strategic complementarity in macroeconomics. In New-Keynesian DSGE models, there is also an interplay between strategic complements and strategic substitutes (see e.g. Woodford, 2003). On the one hand, prices should be high if other firms charge high prices and consumption should be high if other households consume due to the link between consumption and income. On the other hand, through following e.g. a Taylor-type interest rate rule, the central bank introduces strategic substitutability into the model, as an aggressive interest-rate response to undesired inflation raises the real interest rate, thus curbs consumption through the "dynamic IS" equation and hence reduces inflation through the "Phillips curve."[15]
However, even if $f(.)$ is not differentiable, which is usually the case at least over some domain in games with a discrete strategy spaces, one can categorize games into strategic substitutes and complements. (See e.g. Amir, 2005 for a survey.) A game has strategic complementarities if it is "supermodular", implying that agents are incentivized to match other's actions similarly to the continuous space. (See Topkis, 1979, Vives, 1990, and Milgrom and Roberts, 1990 for the theory of supermodular games.) Conversely, a game has strategic substitutes if it

[15] See Assenza et al. (2014) for a more detailed discussion.

is "submodular", implying that agents are incentivized to choose opposite actions to others. Furthermore, according to the axiom of choice in mathematics, an order can be defined over every set (Zermelo, 1904).

3. **Dynamic versus static:** The game can be a one-shot game or a dynamic game. In a dynamic game, the payoff in one round can depend on the decisions in previous rounds. The best-response function in Table 3 introduces a dynamic element by allowing the optimal choice of an individual in a certain period to depend on her belief about future actions and outcomes of other players. The static setup would be a special case with $d=0$. In microeconomic experiments, the repeated interaction of the same static game is the usual research agenda. We will also discuss dynamic games in which state variables are introduced, for example in forecasting in Beauty Contest games where the constant c becomes time-varying.
4. **Simultaneous versus sequential moves:** Games can be played in extensive form (players decide sequentially) or in normal form (players decide simultaneously).

At the end of Section 3, we will discuss other features.

3.2 CONTINUOUS STRATEGY SPACE

Basic Beauty Contest Game

The most basic example for a Beauty Contest game (*guessing game*) has originally been created by Ledoux (1981) and Moulin (1986)[16] and was experimentally tested by Nagel (1995).

The task is that one's choice has to be closest to b times the average:

$$y^i = b\hat{E}^i \frac{1}{N}\sum_{j=1}^{N} y^j \tag{2}$$

of all chosen numbers by the participants, where $y^i \in [0, 100]$ has been restricted to an interval between 0 and 100, imposing integers or real numbers, typically with two decimals. This is a Beauty Contest game with $c=0$ and $f(.) = b \cdot \text{average}$, $d=0$ and $\epsilon_t^i = 0$.

If $0 \le b < 1$ the game has a unique Nash equilibrium $y^{eq}=0$. If $b=1$, any number can constitute a Nash equilibrium, all choosing the same number. With a closed interval choice set and $n>2$[17] the (stable) equilibrium zero, starting at the upper bound can be reached through an iterated elimination of dominated strategies (see also example in the introduction). In open intervals, typically iterated best reply structures lead to the (stable) equilibrium. Finally, if $b>1$ and a closed interval, there are two Nash equilibria, $y^{eq1}=0$ and $y^{eq2} = \textit{upper bound of the interval}$. However, only one of these two equilibria is stable: $y^{eq2}=100$.

[16] See Nagel et al. (2016, 2017) for a historical account of Beauty Contests.

[17] The theoretical case of $n=2$ we discuss in Section 4 together with experimental implementations.

If $b > 0$, the game has strategic complements, as players have an incentive to increase actions when others do so, while if $b < 0$, the game has strategic substitutes, as players need to do the opposite of others. When $b < 0$ and $c = 0$, choices need to include negative numbers. We will alter the number of players, payoff functions, add constants, different information structure about parameters, etc., when discussing experiments related to this game to show how sensitive changes are in terms of equilibrium outcomes, off the equilibrium path structures, and consequentially on behavior. The following games and markets provide part of such changes.

The General Beauty Contest Game

Angeletos and La'O (2010) and Benhabib et al. (2015) show that in a general-equilibrium macroeconomic framework with sentiments (or animal spirits), the equilibrium conditions can be reduced to

$$y_{it} = b\hat{E}_t^i Y_t + \epsilon_t^i \tag{3}$$

where Y_t is the aggregate market outcome, which can, in a simplified fashion, for example be proxied by the average $Y_t = \frac{1}{N}\sum_{j=1}^{N} y_t^j$. $\epsilon_t^i \sim N(c, \sigma^2)$ is an idiosyncratic, private sentiment shock drawn from a distribution which is common knowledge, representing the exogenously given firm specific consumer sentiments.[18] If $b < 0$, the game exhibits strategic substitutes, while for $b > 0$ the game exhibits strategic complements. (d is equal to 0.)

This game is isomorphic to the most general Beauty Contest game discussed in this chapter, $y_t^i = c + b\hat{E}_t^i Y_t + \epsilon_t^i$ with $\epsilon_t^i \sim N(0, \sigma^2)$, being experimentally investigated by Benhabib et al. (2018a, 2018b). If $c = 0$, the Nash equilibrium is playing the private signal $y^{eq} = \epsilon_t^i$, since when all play their signal, the average will be zero and thus the choice for an individual i equals the private signal. If $c > 0$, then the equilibrium becomes $y_t^i = \frac{c}{1-b} + \epsilon_t^i$ and is reached again through an iterated procedure of best reply in an open interval of choices.

Benhabib et al. (2015, 2018a, 2018b) also discuss the case when signals are not precise. Instead, revealed signals are convex combinations of the idiosyncratic signal and a common signal part. In this case players face a signal extraction problem. Since every time a private and a common part is drawn, there can be persistent fluctuations in equilibrium due to this common part. We will discuss an experimental implementation of this theory.

3.2.1 Games of Strategic Substitutes

The Cournot Market

Firm i in the Cournot market (with $N > 2$ firms) chooses the quantity it produces y^i as to maximize its individual profit $\pi^i = py^i - \gamma y^i$ where p is the market price

[18] Sunspots and sentiments are conceptually related to correlated equilibria. Imperfectly correlated signals can create multiplicity of equilibria, as shown by Aumann (1974, 1987) for discrete strategy spaces and by Maskin and Tirole (1987) for continuous strategy spaces.

and $\gamma > 0$ is the marginal cost of producing one unit of the good. Let the (inverse) market demand be $p = a - \psi \sum_{j=1}^{N} y^j$. The first-order condition (and thus the target) is therefore

$$y^i = \psi^{-1}(\gamma + a) - \sum_{j=1}^{N} \hat{E}^i y^j \tag{4}$$

The Cournot game can thus be interpreted as a Beauty Contest game of strategic substitutes with $c = \psi^{-1}(\gamma + a)$, $b = 1$, $d = 0$ and $f(.) = -\sum_{j=1}^{N} \hat{E}^i y^j$ and $\epsilon_t^i = 0$, as this is a standard textbook Cournot market without any exogenous shocks.[19]

Cobweb Market

In Cobweb models (see e.g. Kaldor, 1934; Nerlove, 1958), producers have to produce their commodities first before selling them to buyers. This time lag raises a role for expectations, which can be particularly well illustrated in agricultural markets. Hence, the market can be considered as a price forecasting game so that a producer maximizes profits if her forecast corresponds to the actual price realization: $p_t^{e,i} = p_t$.

Suppose a farmer expects a high price for potatoes. In this case, she would plant more potatoes to harvest at the end of the period. However, if other agents think the same way, the higher supply of potatoes would in fact reduce the price. This is the reason for negative feedback in the Cobweb model so that the reduced form equation is: $p_t = \max\{0, c - b\bar{p}_t^e\}$ where $c, b > 0$ are positive constants, the max operator ensures that prices are non-negative and $\bar{p}_t^e = \frac{1}{N}\sum_{j=1}^{N} p_t^{e,j}$ is the average price expectation in the market for the end of period t.[20] It is easy to see that this is a Beauty Contest game with $d = \epsilon = 0$. For further details see Arifovic and Duffy (2018).

3.2.2 Games of Strategic Complements

Bertrand Market

Let all firms' marginal cost for a product be γ. Firms are submitting prices simultaneously. Assume that the firm with the lowest price obtains the entire market for the product consisting of Q units, so that the individual firm i needs to set the product price p^i as $p^i = \hat{E}^i \min\{\gamma, p^1, \ldots, p^j, \ldots, p^N\} - \eta$ where $\eta > 0$ is small (in order to ensure that player i is below every other player's price). Hence, the Bertrand game can be thought of as a Beauty Contest game with $c = -\eta$, $b = 1$, $d = 0$, and $f(.) = \min\{\gamma, p^1, \ldots, p^j, \ldots, p^N\}$. Note that firms earn higher profits, if they are

[19] Solving for y^i yields $y^i = \frac{1}{2}\psi^{-1}(\gamma + a) - \frac{1}{2}\sum_{j=1, j\neq i}^{N} \hat{E}^i y^j$. Hence, an alternative way of viewing the Cournot game is as a Beauty Contest game with $c = \frac{1}{2}\psi^{-1}(\gamma + a)$, $b = \frac{1}{2}$, and $f(.) = -\frac{1}{2}\sum_{j=1, j\neq i}^{N} \hat{E}^i y_j$, i.e. where players need to decide based on the sum of output of all other players excluding themselves.

[20] Hommes et al. (2007) run a Cobweb experiment with a non-linear aggregation function due to non-linear aggregate supply given by a tanh-function.

able to set higher prices p^i, as firms' profit is given by $\pi^i = p^i q^i - \gamma q^i$. The unique (inefficient) Nash equilibrium is, however, $p^i = p^1 = \cdots = p^N = p^{eq} = \gamma$.

Auctions

In a first-price sealed bid auction, all participants simultaneously submit their bids in a way that no agent knows the bid of other participants. In a buyer auction, the individual with the highest bid wins.[21] Hence, every participant chooses a bid y^i so that $y^i = \hat{E}^i \max\{\gamma_i, y^1, \ldots, y^j, \ldots, y^N\}$, where $\gamma_i \geq 0$ may be the seller's (privately known) redemption value. It is easy to see that this is a Beauty Contest game with $c = 0$, $b = 1$, $d = \epsilon = 0$. The γ_i can be thought of as being given by a constructed player called "nature". In Bayesian Nash equilibrium each player chooses $(n-1)/n^* \gamma_i$.

Asset Markets

Campbell et al. (1997) and Brock and Hommes (1998) outline a standard mean-variance asset pricing model with heterogeneous expectations and two assets: a riskless asset with perfectly elastic supply and a risky asset with fixed, perfectly inelastic supply. One very simple design to introduce asset markets as a laboratory setup is a "learning to forecast" design in which subjects are only paid for forecasting and based on subjects' forecast the computer optimizes for them. A particular subject j is then paid according to a distance function such as $U_t^j = A - B(\hat{E}_t^j p_{t+1} - p_{t+1})^2$, i.e. how close her forecast for period $t+1$, $\hat{E}_t^j p_{t+1}$, is to the realized price in $t+1$, p_{t+1}.

Campbell et al. (1997) and Brock and Hommes (1998) both use a market clearing asset pricing model (i.e. with a Walrasian auctioneer setting the equilibrium price). In terms of learning-to-forecast experiments, this leads to a two-period ahead forecasting game (as in Hommes et al., 2005). For the positive feedback asset market Heemeijer et al. (2009) implement a learning-to-forecast asset market, using a market maker price adjustment rule. The reason for this is that they want to compare positive versus negative feedback markets. An asset market with a market maker reduces to a one-period ahead forecasting game with the same timing as in the classical cobweb model with negative feedback.

Under market clearing, the price, being the target in learning-to-forecast experiments, can be expressed by $p_t = \frac{1}{1+r}(\bar{p}^e_{t+1} + \bar{y} + u_t)$ where $\bar{p}^e_{t+1} = \frac{1}{N}\sum_{j=1}^N \hat{E}_t^j p_{t+1}$ is the average belief in the market at period t about the price in period $t+1$, r is the interest rate, $\bar{y}$ is the mean dividend and u_t captures supply and demand shocks in the market.

This is a Beauty Contest game of strategic complements with $c = \frac{\bar{y}}{1+r}$, $b = 0$, $d = \frac{1}{1+r}$, $f(.)$ being the average and the shock being the same for all individuals:

[21] Another type of auction that is often used in practice is the second-price sealed bid auction, in which the individual with the lowest bid wins but receives the second-lowest bid. In a second price auction, bidding one's value is a weakly dominant strategy.

$\epsilon_t^i = \epsilon_t = \frac{u_t}{1+r}$. The current price p_t is given as a temporary equilibrium, being a function of the average belief of the price in the next period, $t+1$.

Another model of price formation is a market maker, like a Walrasian auctioneer, calling out prices and aggregating the asset demand for a given price (Beja and Goldman, 2015). Under this mechanism, the price is determined by a linear function $p_t = c + b\bar{p}_t^e + u_t$, where $\bar{p}_t^e = \frac{1}{N}\sum_{j=1}^N \hat{E}_t^j p_t$ is the average forecast in the market at period t about the current price in t. (See e.g. Heemeijer et al., 2009.) It is easy to see that this is also a Beauty Contest game. For further details see Arifovic and Duffy (2018).

New-Keynesian Models

A heterogeneous expectations version of the 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

$$y_t = \bar{y}_{t+1}^e - \sigma(i_t - \bar{\pi}_{t+1}^e - \rho) + g_t \tag{5}$$

$$\pi_t = \kappa y_t + \beta\bar{\pi}_{t+1}^e + u_t \tag{6}$$

$$i_t = \rho + \phi_\pi(\pi_t - \pi) \tag{7}$$

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.

Similarly to the asset market case, we consider a learning-to-forecast game, in which subjects are incentivized to forecast inflation and output gap, being generated as temporary equilibrium where inflation and output gap of period t depend on $\bar{\pi}_{t+1}^e$ and $\bar{y}_{t+1}^e$, respectively, which are the average of all subjects' forecasts for period $t+1$. An individual subject j is then paid according to a distance function such as $U_{t+1}^j = A - B(\hat{E}_t^j \pi_{t+1} - \pi_{t+1})^2$, i.e. how close her forecast for period $t+1$ (given in period t), $\hat{E}_t^j \pi_{t+1}$, is to the realized inflation in $t+1$, π_{t+1}, and similarly for the output gap. It is easy to see that this model is a special case contained in the Beauty Contest definition, since we allowed the variable of interest to be a vector. For the sake of simplicity, however, we describe the New-Keynesian model as a univariate forecasting game of inflation at time t, π_t.

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,[22] obtained by using Eq. (6):

$$\bar{y}_{t+1}^e = \kappa^{-1}(1-\beta)\pi \tag{8}$$

[22]To keep the task for subjects simple, laboratory experiments have used different ad-hoc assumptions. Pfajfar and Zakelj (2014) use naive expectations about the output gap. Assenza et al. (2014), Arifovic and Petersen (2017), and Mauersberger (2016) use forecasts generated by subjects for both inflation and output gap.

Substituting (8) into (5), the system becomes

$$y_t = \kappa^{-1}(1-\beta)\pi + \sigma\phi_\pi \pi - \sigma\phi_\pi \pi_t - \sigma\bar{\pi}^e_{t+1} + \kappa g_t \tag{9}$$

$$\pi_t = \kappa y_t + \beta\bar{\pi}^e_{t+1} + u_t \tag{10}$$

By inserting (9) into (10) and rearranging one obtains the target that subjects need to forecast at time $t-1$:

$$\pi_t = c + d\bar{\pi}^e_{t+1} + \upsilon_t \tag{11}$$

with $c \equiv \frac{(1-\beta)+\kappa\sigma\phi_\pi}{1+\kappa\sigma\phi_\pi}\pi$ and $d \equiv \frac{\beta-\kappa\sigma}{1+\kappa\sigma\phi_\pi}$ and υ_t being a composite shock at time t. It is easy to see that the New-Keynesian model can be considered to be a Beauty Contest game with $c > 0$, $d > 0$, $b = 0$, $f(.) = \bar{\pi}^e_{t+1}$, and a common random term $\epsilon^i_t = \upsilon_t$ being the same for every individual i. For further details see Arifovic and Duffy (2018).

3.2.3 Public and Private Information

Unknown Constant Fundamental

Morris and Shin (2002) consider the following Beauty Contest game:

$$y^i = (1-r)\hat{E}^i \frac{1}{N}\sum_{j=1}^{N} y^j + r\hat{E}^i c \tag{12}$$

where $0 \le r \le 1$. Agents do not have full knowledge of c. However, agents observe a public signal of c, denoted as $z = c + \eta$ where $\eta \sim N(0, \sigma^2_\eta)$ is a Gaussian random shock, and also a private signal of c, being denoted as $x^i = c + \xi^i$, where $\xi^i \sim N(0, \sigma^2_\xi)$ is a Gaussian random shock. Hence, agents face a *signal-extraction problem*. In equilibrium, agents play a convex combination of these two signals, denoted as $y^i = \theta x^i + (1-\theta)z$, where $\theta = \frac{(1-r)\sigma^2_\eta}{(1-r)\sigma^2_\eta + \sigma^2_\xi}$. While Morris and Shin (2002) consider the case of strategic complements, one could also consider a game of this type with strategic substitutes, which would be obtained if $r > 1$ or more generally as $y^i = b\hat{E}^i \frac{1}{N}\sum_{j=1}^{N} y^j + \lambda\hat{E}^i c$.

3.2.4 Other Games

Public Goods Games

The public good game we mentioned in Section 2 is neither attributed to strategic complements nor substitutes since other player's decisions do not enter an individual i's best response. There is a dominant strategy both in public good games with a boundary solution and public good games with an interior solution. Yet other player's decisions are also payoff-relevant. Consider N individuals that are asked to contribute an amount y^i_t to a joint project (that is non-excludable and non-rivalrous in consumption) at time t from their endowment e^i_t. A boundary solution is obtained, if, by contributing y^i_t, individual i obtains the payoff $u^i_t = e^i_t - y^i_t + m\sum_{j=1}^{N} y^j_t$.

It is easy to see that individual i's dominant strategy is $y_t^i = 0$. An interior solution is obtained, if, individual i receives a payoff function of for example $u_t^i = e_t^i + n \cdot y_t^i - y_t^{i,2} + m \sum_{j=1}^{N} y_t^j$ (similar to Keser, 1996) with a dominant strategy $y_t^i = \frac{m+n}{2}$ for all players i. However, despite the absence of strategic substitutes or complements, this can be seen as a Beauty Contest game with $c = 0$ for the boundary-solution case and $c = \frac{m+n}{2}$ for the interior-solution case with $b = d = \epsilon = 0$.

3.3 DISCRETE STRATEGY SPACE AND MULTIPLICITY

As mentioned before, classifying games into games of strategic substitutes and complements requires an ordered strategy space. An order obviously exists in games with continuous strategy spaces, because numbers can be ranked from "lowest" to "highest." One can also categorize games with discrete strategy space into games of strategic substitutes and complements. (See e.g. Amir, 2005.) A game has strategic complementarities if it is "supermodular", implying that agents are incentivized to match other's actions similarly to the continuous space. It is easy to see that a prisoner's dilemma (PD) game can be considered as a public good game with two strategies, say 0 and 100, and two players, reduced from a continuous [0, 100]-interval. Similarly, a stag-hunt game corresponds to a minimum effort game with a reduced strategy space.

A contrary case to the PD is the harmony game. In the harmony game, the dominant action for all players is also Pareto-efficient, while this is not the case in the PD.

3.3.1 Strategic Complements

The Original Keynesian Beauty Metaphor

Keynes (1936) outlined the following game: Each individual i has to make a choice y^i out of a finite, discrete set (A, B) so that in a contest with N individuals:

$$y^i = \hat{E}^i \, \mathrm{Mode}\{y^1, y^2, \ldots, y^i, \ldots, y^j, \ldots, y^N\} \tag{13}$$

where $\hat{E}^i$ denotes the subjective belief of individual i, accounting for the information friction that individual i may not know the choices of the other individuals $j = 1, \ldots, N$. An equilibrium strategy combination is given by any choice, provided that all players make this choice. This is a Beauty Contest game with $f(.)$ being the model. One can also formulate this game for the continuous case, in which case it is easy to see that it would be a Beauty Contest game with $b = 1$, $c = 0$, $d = 0$. However, this game does not induce higher order beliefs despite the fact that Keynes inserts higher order reasoning mentioned in the introduction into this game. If a player believes that face A is chosen by most others, then she herself has also to choose face A. Thus, there is no higher order belief involved. This property holds for all games discussed in this subsection unless otherwise stated.

Table 4 Stag-hunt game

	Stag	Hare
Stag	2, 2	0, 1
Hare	1, 0	1, 1

Table 5 Van Huyck et al. (1991): Median effort game

		Median value of X chosen						
		7	6	5	4	3	2	1
Your choice of X	7	1.30	1.15	0.90	0.55	0.10	−0.45	−1.10
	6	1.25	1.20	1.05	0.80	0.45	0.00	−0.55
	5	1.10	1.15	1.10	0.95	0.70	0.35	−0.10
	4	0.85	1.00	1.05	1.00	0.85	0.60	0.25
	3	0.50	0.75	0.90	0.95	0.90	0.75	0.50
	2	0.05	0.40	0.65	0.80	0.85	0.80	0.65
	1	−0.50	−0.05	0.30	0.55	0.70	0.75	0.70

Stag-Hunt Game

Consider for example a 2×2 *stag-hunt game*, whose payoff matrix is depicted in Table 4. In this game payoffs are high, if both players either play stag or hare, but lose out if one player plays stag and the other hare. Hence, players are well-off, if they choose the same actions, i.e. the mode of $\{y^1, y^2\}$ being unique, while they lose out if they choose different actions, i.e. $\{y^1, y^2\}$ having two modes. This is a Beauty Contest game in the discrete strategy space with $f(.)$ being the mode.

Minimum/Median Effort Game

Similarly to the stag-hunt game whose two equilibria can be Pareto-ranked, Van Huyck et al. (1990,1991) introduce a game into the laboratory where players choose a number between 1 and 7 with the interpretation that a higher number represents higher effort. Each number is associated with a fixed payoff depending on a player's own choice and the minimum or median choice of all other players. An example of such a payoff structure is given in Table 5. In an equilibrium all have to choose the same number, with all choosing 7 being the Pareto optimal equilibrium and all choosing 1 being the risk dominant equilibrium.

Global Games

A latent assumption in the Keynesian Beauty Contest game is common knowledge about the economic fundamentals. Morris and Shin (2001) purport that this symmetry in knowledge is the source of indeterminacy in equilibrium. Consequently, they show that the presence of some uncertainty about the fundamentals can eliminate the multiplicity in equilibria. Carlsson and van Damme (1993), who introduced the term "global games", have shown that small perturbations in the payoff matrices or in the

Table 6 Global games (entries represent players' payoffs)

	Status quo (if $M < f(\theta)$)	Alternative (if $M \geq f(\theta)$)
Attack	$-\lambda$	ψ
Not attack	0	0

information in generalized 2×2 games with multiple equilibria can generate a unique rationalizable equilibrium.

Angeletos et al. (2007) interpret global games as scenarios of regime switch. There are two possible regimes: "status quo" and "alternative" (Table 6). Like in the original Beauty Contest game, there is a discrete set of actions $K = (0, 1)$. Each player can choose one action being preferred in the status quo regime, "not attack" ($k^i = 0$), and one action being preferable in the alternative regime, "attack" ($k^i = 1$). The payoff from not attacking can be normalized to zero and the payoff from attacking can be $\psi > 0$ if the status quo is given up and $-\lambda < 0$ if retained. This implies that individuals should optimally attack, if they expect the status quo to be abandoned. Whether the status quo is abandoned or not depends on whether the aggregate mass of attackers M exceeds a certain critical mass $f(\theta)$, which is a monotonically decreasing function of θ. If less than $f(\theta)$ agents choose attack, all choosing attack earn $-\lambda$, if more than $f(\theta)$ agents attack, all those who attack earn ψ. If θ is common knowledge, there are two equilibrium outcomes: everyone attacking and nobody attacking. A unique equilibrium is only attained, if there is heterogeneous information in the form of a noisy signal about the fundamental. If a private signal of the form $x^i = \theta + \sigma_\epsilon \epsilon^i$ is drawn where ϵ follows a standard normal distribution, Morris and Shin (1998, 2001) show that the size of the attack K is monotonically decreasing in θ and the regime switch occurs if and only if $\theta < \theta^*$, where $\theta^* = 1 - \frac{\psi}{\lambda}$. If there is in addition a public signal of the form $z = \theta + \sigma_\xi \xi$ with ξ following a standard normal distribution, Morris and Shin (2002) show that a unique equilibrium requires $\sigma_\xi \epsilon \leq (2\pi)^{1/2} \sigma^2 \xi$. The survey by Angeletos and Lian (2016) mentions numerous applications for global games: currency crises (Obstfeld, 1996), bank runs (Goldstein and Pauzner, 2004; Rochet and Vives, 2004; Corsetti et al., 2006), debt runs (He and Xiong, 2012), the role of credit rating agencies (Holden et al., 2014), business cycles (Schaal and Tascherau-Dumouchel, 2015), and the particular role of prices as public signals (Angeletos and Werning, 2006). [23]

Schelling (1960)

Schelling (1960) purports that a game of coordination is at the same time a game of conflict, e.g. arranging a time for a meeting involves coordination on a time and avoiding possible scheduling conflicts. These games can be considered to be equivalent to the Keynesian Beauty Contest game. If coordination is supposed to be

[23] Global games have also been formulated for strategic substitutes. See e.g. Karp et al. (2007).

tacit, according to Schelling, whether individuals converge to an outcome depends on whether expectations coordinate on a focal point.

He also understood segregation as an outcome of coordination: supposing individuals want to have a certain number of neighbors of the same ethnicity, then, for instance, one black household moving away may induce other black households to move away. More formally, an individual has to make a choice y^i of a location $x \in X$, with the best response

$$y^i = x^* \text{ s.t. at least } h > 0 \text{ individuals choose } x^* \tag{14}$$

Schelling conducted an early experiment about this by playing with different types of pennies on the chessboard that needed to be surrounded by other pennies (Dixit, 2006). This is a Beauty Contest game with $c = 0$, $b = 1$, $d = 0$ and $f(.)$ being the choice made by at least h individuals.

11–20 Game

Arad and Rubinstein (2012a) propose a two-player game in which each player chooses an integer in a fixed interval, say 11 to 20. The player receives her choice as a payoff and a large bonus (e.g. a payoff of 20 in addition) if she is exactly one below her opponent's choice. In Arad and Rubinstein's original game, there is no Nash equilibrium in pure strategies but only a mixed strategy equilibrium corresponding to playing 15, 16, 17, 18, 19 and 20 with probabilities 25%, 25%, 20%, 15%, 10%, 5% respectively.

Alaoui and Penta (2016a, 2016b) have thus slightly modified the game, also giving the bonus in the case of a tie, which renders the equilibrium unique at 11. The game is therefore also dominance solvable as the BC-game. Fragiadakis et al. (2017) have modified the game by giving individuals the bonus if they are $r = 3$ (in some games $r = 4$) below the opponent's choice and an additional smaller bonus in case of ties, which allows a more distinctive analysis for belief elicitation in cognitive hierarchy models. This game has multiple equilibria, e.g., both choosing either 11, 12, or 13, if $r = 3$, as undercutting is not possible for these numbers.

3.3.2 Strategic Substitutes

Hawk–Dove Game

An example of strategic substitutes would be a game with the individual best response being:

$$y^i = \text{Choice least made in } \ldots \{y^1, y^2, \ldots, y^i, \ldots, y^j, \ldots, y^N\} \tag{15}$$

where y^i may again be a choice out of a finite, discrete set $\{A, B, \ldots\}$.

Eq. (15) can be considered to be an accurate description of payoffs the hawk–dove game, in which whose payoff-matrix is depicted in Table 7.

If player 1 (2) plays dove so that $y^1 = D$ and player 2 (1) plays hawk $y^2 = H$, then both choices satisfy $y^i = \text{Choice least made in } \ldots \{y^1, y^2\}$, which are equilibria. This is not the case if both play either hawk or dove.

Table 7 Hawk–dove game

	Hawk (H)	Dove (D)
Hawk (H)	−1, −1	1, 0
Dove (D)	0, 1	1/2, 1/2

Ultimatum Game

$N = 2$ players need to split a pie of size $c > 0$. Both announce y^1, y^2 (possibly at different point in times) how much of the pie they claim. A requirement is that $y^1 + y^2 \leq c$, as the allocation is otherwise not feasible, in which case every player receives a low default payoff. Each player i's best-response given by $y^i = c - y^{-i}$. Hence, if players move simultaneously, the ultimatum game can be considered a Beauty Contest game with strategic substitutes where $c > 0$, $b = -1$, $d = 0$, $f(.) = y^{-i}$ trivially being the choice of the other player and no random element ($\epsilon = 0$). All combinations of offers and demands with the requirement $y^i + y^{-i} = c$ are equilibria. If played sequentially,[24] with one player making the offer and one accepting or rejecting it, the lowest positive offer accepted is the subgame perfect equilibrium. However, if there is competition between proposers (responders), the game is of strategic complements, as the competitors have to increase their offers (decrease their demands) if the others also do it. Then the equilibrium with the highest possible offer (lowest demand) is unique and subgame perfect.

3.3.3 Matching

The fact that in the Keynesian Beauty Contest game one must *match* the action of the average reveals the link to a widely used concept in economics: matching. The Keynesian Beauty Contest game is an example of *one-to-many matching*. Other forms of matching are *many-to-one matching*, for example one university admitting many students or just *one-to-one matching*, meaning that two agents pair up. Examples would be employee-employer hirings, school choice, marriage markets, kidney exchange etc.

One distinguishes between *positive assortative matching* on the one hand, which resembles strategic complementarities and means that agents match with similar agents. The opposite would be *negative assortative matching*, which resembles strategic substitutability and means that agents match with agents of dissimilar or opposite characteristics. Heterogeneity is the root of matching problems, including the Keynesian Beauty Contest game, as in the presence of homogeneity any matching problem would become trivial. See Roth et al. (1990) for a more detailed treatment of two-sided matching markets.

[24] In dynamic coordination games, there are good reasons to believe that the players move asynchronously. For example, either players alternate in moves or because opportunities to switch actions may arrive randomly to different players. The question whether such an asynchronous nature of moves would help or hinder coordination has been investigated by a large theoretical literature, including for instance Lagunoff and Matsui (1997), Gale (1995), Morris (2014), Matsuyama (1991), and Matsui and Matsuyama (1995).

Table 8 Matching pennies (Hide and Seek)

	Head (H)	**Tail (T)**
Head (H)	1, −1	−1, 1
Tail (T)	−1, 1	1, −1

3.3.4 Strategic Heterogeneity

Another possibility is a game in which some agents face strategic complements, i.e. need to match other agent's actions, while others face strategic substitutes, i.e. need to do the opposite of others' actions. This has been labeled as strategic heterogeneity (Monaco and Sabarwal, 2016).

Matching Pennies

A well-known textbook example is matching pennies (or hide and seek). This is a game with two players (A and B). Every player has a coin and secretly has to turn it on the head or tail side. They then simultaneously reveal their choices. If the coins match (both showing either head or tail), player A gains the coin of B. If the coins do not match, B gains the coin of A. This can easily be illustrated by a payoff matrix shown in Table 8.

It is easy to see that this game has a unique mixed strategy Nash equilibrium in which both players play each action with probability 50%.

While this seems a stylized example, several real world applications have been found. For example, some sport games are isomorphic to matching penny games (or "generalized matching penny games" (Goeree et al., 2003) with perturbed payoff structures). In soccer penalty kicks, both the kicker and the goalkeeper have two actions – left and right – and one of them has to match the other's action, while the other one has to do the opposite of the opponent.

Fashion Cycles

A somewhat similar case to matching pennies is the evolution of fashion. Matsuyama (1992) studies fashion cycles in a simple model with two actions – red and blue – and two types: conformists who like dressing like others and non-conformists who like dressing differently from others. A vital difference to matching pennies is the path dependence in the dynamic version of the game, since agents switch between red and blue over time depending on both their type and on what the majority chooses. Matsuyama shows that, depending on the share of the types and the matching patterns, this can lead to a stable limit cycle in which Nonconformists become fashion trend-setters switching their actions periodically, while Conformists follow with some inertia.

The Colonel Blotto Game

The Colonel Blotto game can be framed as every player being the commander of an army, having a fixed number of troops. At the beginning of each round, every player has to distribute her number of troops across a fixed number N of battlefields without observing the opponents' choices. Once every player has decided upon her allocation of troops, the battle begins according to the following rules: on every battlefield, the player who has deployed the most troops will win. The player(s) having won the most battlefields is (are) the winner(s) of that round. The Colonel Blotto game as a static game shares some similarity with the fashion cycles model, since there are both motives to match the decisions of the majority, i.e. put many (few) troops on battlefields if opponents put many (few) troops there, but also do the opposite to the majority, i.e. avoid investing troops in a field in which many players put many troops.

3.4 OTHER DIMENSIONS

There are certainly many other dimensions, which may contain further sources of heterogeneity. We distinguish between dimensions that are induced by the game structure on the one hand and dimensions that are related to the experimental implementation.

3.4.1 Game Structure

- **Number of human players:** Players can play against other human players or against machines or nature. There can be one player, which would constitute a game against nature, typically different from playing against humans, as it does not involve e.g. forming higher order beliefs. We will show examples with various numbers of players. The two-player case, which is standard in many textbook games, is often very different both in theory and in the observed outcomes from the case with more than two players (Sutan and Willinger, 2009). It crucially depends on the context whether the number of players empirically makes a difference: Bao et al. (2016) use experimental groups of more than 20 subjects for a learning-to-forecast experiment in an asset market setup, finding that behavior is similar to the experiments with small groups. In contrast to that, Arifovic et al. (2018) investigate whether sunspot announcements have different effects on small groups (of 10 persons) as opposed to large groups (of 80–90 persons) in a bank-run setting, where people decide whether or not to withdraw their money from the bank. They find that none of their large groups coordinate on a random sunspot signal, while small groups sometimes do. (See chapter by Arifovic and Duffy, 2018.)
- **Role of information:** There is often a distinction between imperfect and incomplete information. What is meant by these terms may vary slightly depending on the context. In game theory, imperfect information means that agents are not informed about the actions chosen by other players. Yet, they know the "type" of other players, their possible strategies/actions and their preferences or payoffs. Incomplete information, on the other hand, means that players do not know the "type" of the other players, so that they may lack information about their strategy profile or their payoff. Angeletos and Lian (2016) define imperfect information as agents being uncertain about a fundamental value, e.g. merely receive a signal about the fundamental value, while incomplete information is

defined as agents not being informed about the whole distribution of information across the population, e.g. they do not know which signals other agents receive.

Another debate is whether subjects should be given the functional form of the best response function or the data-generating process. (See Hommes, 2011 for a detailed survey.) While the functional forms of the best responses are often given to the subjects in experimental games, it is often argued (see e.g. Hommes, 2011) that the laws of markets are unknown by market participants so that in experimental finance or macroeconomics, subjects are usually only provided with qualitative information (e.g. Hommes et al., 2005) or no information at all (e.g. Adam, 2007).

- **Calibration:** By characterizing games as complements or substitutes, we made distinctions only according to whether $f(.)$ is increasing or decreasing in the actions of others. However, another important consideration is how much $f(.)$ increases or decreases in the actions of others, which may depend on exogenous parameters that need to be calibrated. For example, we will exemplify below that it makes a difference whether in a simple best response function of $y_t^i = b \cdot \frac{1}{N} \sum_{j=1}^{N} y_t^j$ the coefficient b is chosen to be 0.67 or 0.95.
- **Payoff function:** Players may be rewarded according to a tournament payoff so that one or several winners of the game are paid a fixed prize. As opposed to that, all players can alternatively be rewarded according to a (quadratic) distance function between own choice and the target choice, the outcome of the games. Typically, tournament payoffs produce more outliers, as incentives are too weak or non-existing for most of the players. Deviations are therefore more likely (Sonnemans and Tuinstra, 2010). Theoretical properties can also be very different between these two different payoff schemes which we will discuss in more detail in Section 4.

  A particularly important feature is flatness of the payoff function (around the equilibrium). With a too flat payoff function, suboptimal behavior may not be punished sufficiently, because subjects may face too weak incentives to think about making better decisions.
- **Payoff heterogeneity:** Players can be subdivided into different groups that interact with each other but that are assigned different tasks or that have different payoff functions. For example, if a general equilibrium model is implemented into the laboratory, some players may play firms and others households.
- **Order statistics:** As implied by the function $f(.)$ in Table 3, there are different ways how the action of other players can be aggregated. Hence, the payoff can depend on the sum, mean or median of the (other) participants' actions. The median obviously excludes single outliers in the interaction. However, more intricate dependencies such as network structures can be possible where only some neighbors determine a player's payoff.
- **Type of interaction between several players:** The experimenter has some degree of freedom on how persons interact in an experiment. People can interact with each other directly or anonymously. Furthermore, the experimenter may choose

to divide the subjects into subgroups that may play against each other, e.g. in a network with different neighborhoods. In particular, in repeated games the matching mechanism may matter, as players can be randomly matched in each round (repeated one-shot games) or keep playing against the same person (supergames).

3.4.2 Experimental Implementation

- **Sociological dimensions:** Sometimes the researcher's focus is on sociological dimensions such as gender, ethnic group or age. We have already discussed in Section 2 how results depend on culture and gender.
- **Methodology:** The researcher may want to address whether the methodology used makes a difference. For example, he may wonder whether it makes a difference that the data is collected in the laboratory or from the field.
- **Cognitive dimensions:** The experimenter may want to address whether cognitive dimensions, such as risk preferences, IQ, emotions etc., matter for the outcome. This often requires some test, for instance an IQ test, before the game or some form of priming, e.g. inducing a certain type of emotion.
- **Framing:** Game-theoretic experiments and Beauty Contest games (see e.g. Nagel, 1995; Duffy and Nagel, 1997; Ho et al., 1998) usually have an abstract framing without describing any concrete situation. Other games have an economic framing, explicitly communicating to participants that they are for example forecasters in a virtual macroeconomy, investors in a financial market or firms choosing production in an oligopoly market. A concrete framing is often chosen in more sophisticated games, as an abstract framing often further increases the level of complication. Game theoretically, the language or the contexts that are presented should not induce differences in optimal behavior. However, humans might react very differently to small changes in the presentation of the same mathematical models.
- **Structural knowledge:** In microeconomic experiments, most aspects of the game structure, such as parameters of the game, as e.g. the b-parameter, number of players, are typically communicated to the subjects in advance. There might be a historical reason behind this informational setup. At the outset of the experimental economics literature, the main hypotheses tested in the laboratory were whether subjects play the equilibrium. Initial tests were conducted in their strongest form, i.e. by confronting players with the same or similar information as a theorist would need to analyze the game. However, the rationality requirements or more so common knowledge of rationality are outside the control of the experimenter. Furthermore, assuming the level of knowledge about the structure of the model is not always realistic and in fact there are strands of the economic literature, such as the macroeconomic learning literature (Marcet and Sargent, 1989; Sargent, 1993; Evans and Honkapohja, 2001), where the underlying environment is unknown to the decision-maker in the model. For an experimental implementation, this can mean that the parameters are unknown, or that the relationship between variables is only given in a qualitative way, e.g. whether one variable increases when another variable increases. This is a possible motivation to also analyze be-

havior under less information than a theorist in the abstract might have and to test whether agents are able to learn the rational expectations equilibrium over time. In this chapter subjects will always know the structure of the model, like the theorist. Arifovic and Duffy (2018) discuss experiments with less knowledge. This also gives rise to experiments which are much more complicated than in the microeconomic world (see e.g. Adam, 2007) or only described qualitatively (see e.g. Nagel and Vriend, 1999; Lei and Noussair, 2002; Hommes et al., 2005, 2008; Heemeijer et al., 2009; Bao et al., 2012; Pfajfar and Zakelj, 2014; Mauersberger, 2016; Sonnemans and Tuinstra, 2010). Such experiments can be considered to test the theoretical predictions of a large adaptive learning literature and bounded rationality literature in finance and macroeconomics (see e.g. Sargent, 1993; Evans and Honkapohja, 2001), which hypothesizes that agents do not know the coefficients of the underlying model equations, but they need to learn them from past data acting like econometricians, or their perceived law of motion is misspecified.

4 BEHAVIORAL REGULARITIES IN BC-EXPERIMENTS AND LEVEL-K

In Section 3 we show that the generalization of the Beauty Contest game provides a framework for many seemingly different models. In this section we introduce a behavioral model that is useful for encompassing or structuring behavior in many different (BC)-experiments, with the idea to suggest a framework for boundedly rational behavior. The main focus in the BC experimental literature is about the study of cognitive ability by subjects to understand the rules of the game, and the ability to outguess or predict the behavior of others, called strategic intelligence. Based on the behavioral patterns in the first BC-experiment implemented in the lab, Nagel (1993, 1995) introduced a level of reasoning model, later called level k, to describe these behavioral regularities. She provides empirical evidence and visualization for the simplest formulation of level k by using an appropriate, simple experimental game.

Over the past 25 years it could be shown that this behavioral model forms a framework for classifying and structuring heterogeneous behavior in the lab and in the field across many different situations. First of all it describes how bounded rational players interact. Also, in the spirit of von Neumann's question posed at the beginning of this chapter, the model suggests how homo economicus should interact with nonrational or boundedly rational players, or react to other external circumstances. More generally, what kind of heuristics do subjects formulate? Do structural changes of the game that may or may not shift the equilibrium also alter behavior? This section is by no means a complete survey of BC games, nor does it cover all experimental or theoretical papers discussing level-k applications. However, we hope to discuss the important issues about a framework of behavioral patterns which will raise new interesting questions.

Guided by the game or market models discussed in Section 3, we organize the variations of the Beauty Contest game that differ along several dimensions: the magnitudes, and the signs of the game's parameters, i.e. whether the game is characterized by strategic substitutes or by strategic complements; the information structure given to the subjects; the payoff function determining subjects' rewards; unique versus multiple equilibria, and the presence of exogenous stochastic, payoff-(ir)relevant elements.

Yet, we also take into account factors independent of the mathematical formulations, as for example the characteristics of the subject pool. Do people react differently when they are informed about the level of sophistication of other subjects? Since behavioral heterogeneity in experiments is often due to cognitive differences between subjects, typically equilibrium play is initially not observed in the laboratory, even if the equilibrium is unique and payoff-maximizing for all players, and thus socially optimal. We show different layers of heterogeneity. In the first period and over time on the one hand and across subjects on the other hand. The quest is also how to design the game to obtain efficiency-enhancing or optimal choices instantaneously. We discuss the simple structures of the BC games. The experimental literature on more sophisticated BC games and markets such as asset markets, Cobweb, and New-Keynesian models, mentioned in Section 3, are discussed by Arifovic and Duffy (2018).

4.1 THE BASIC BC EXPERIMENT

The main features of most of the games we discuss is the aggregation of behavior as an average or another order statistic such as the median, with $n \geq 2$ or $n = 2$ players. Nagel (1995) introduces the basic BC game ($y^i = b\hat{E}^i \frac{1}{N}\sum_{j=1}^{N} y^j$ and tournament payoffs) to the experimental literature. A large number of players, between 15 and 18, interacts for 4 periods within the same group. In this experiment, there are two salient results: first-period behavior is far away from the equilibrium, all choosing zero, but behavior slowly converges over time towards zero. To find regularities, the pen-and-paper experiment consists of BC games with the parameters $b = 1/2$, $2/3$, and $4/3$. Each group is exposed only to one parameter value and every subject was only allowed to participate only in one group.

4.1.1 First Period Choices

Fig. 7 shows first period behavior of different subject pools with $b = 2/3$, including Nagel's laboratory data in the first row, first graph. Choices are highly dispersed but with clear patterns, spikes near and at 22, 33, and 50. There are also choices near 67, because some subjects may calculate $100 \cdot (2/3)$. Out of all subjects only few choose zero. We discuss the entire figure in Section 4.1.4.

4.1.2 The Level-k Model and Related Models

The Basic Level-k Model

Nagel (1995) conceptualized the basic level-k model (then called "steps of reasoning model") based on her own experience as a subject in a two-period 2/3-BC-classroom experiment (see Nagel et al., 2017). She participated in this classroom experiment

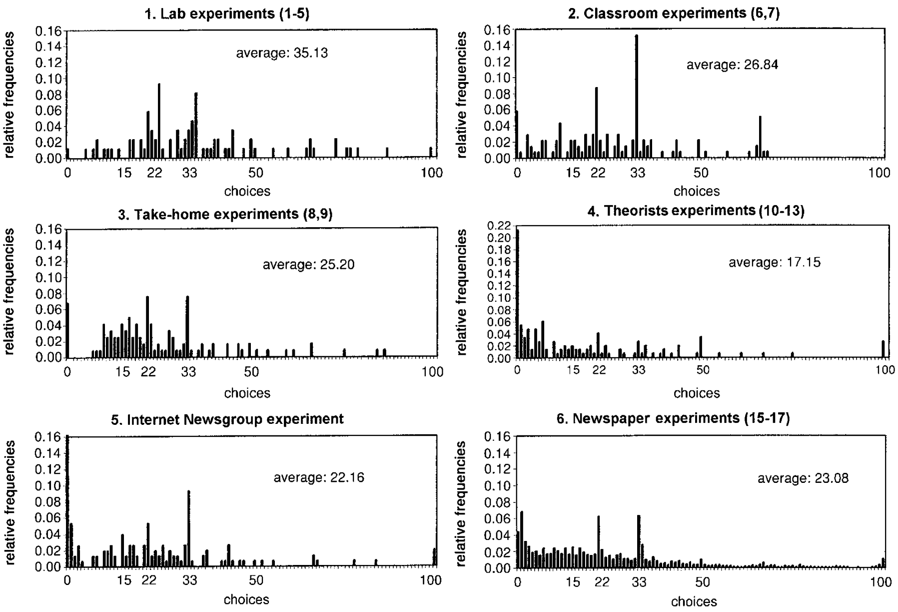

FIGURE 7

Bosch et al. (2002): Relative frequency of choices and pooled averages within the six different subject pools (b = 2/3). Group sizes range from 15 to 18 in the lab, 80–100 in the classroom, 20–30 for theorists, 40 in the newsgroup, and 2000–3000 in the newspaper experiments. (The number in parentheses indicate the different session numbers within each subject pool.) For the complete dataset, see Nagel et al. (2018).

as a graduate student at the London School of Economics in 1991 in a game theory lecture by Roger Guesnerie. Support for level-k has been found in the aforementioned experimental datasets and also in the written comments requested from the subjects.[25]

The model consists of a reference point, called level 0 and (finite) iterated best replies. Within the behavioral interactive realm this rule might also have a flavor of a generative principle, which should be studied in the future. It is assumed that all (naive) players in a Beauty Contest game choose randomly, with an average of 50 in the interval [0, 100], for insufficient reasoning. A level-1 (L1) player anticipates this and chooses the best response, $50 \cdot 2/3 = 33.33$, if $b = 2/3$. A level-2 (L2) player anticipates a level-1 player's choice and chooses $22.22 = 50 \cdot (2/3)^2$. Using the same logic, a level-k (Lk) player reacts to level k − 1 (Lk-1) player with $50 \cdot (2/3)^k$. That way, a player who believes that all players will come to the same conclusion and iterates infinitely, will reach the equilibrium zero. Given that equilib-

[25]This idea of players with different degrees of sophistication has, for example, theoretically been explored independently by Stahl (1993), calling the levels of sophistication "smart$_n$". See Banerjee et al. (1996) for a survey on evolutionary game theory from a purely theoretical perspective.

rium play typically does not win, a player who chooses it, is likely to suffer from a curse of knowledge of the mathematical solution, being ignorant of a winning strategy against boundedly rational subjects (see Camerer et al., 1989 who discuss this curse in the context of information asymmetries). Of course, one cannot expect that subjects make such degenerate choices. For instance, they might just make an approximation to $50 \cdot 2/3$ such as 30. Thus, Nagel allowed for noise by constructing intervals around such theoretical numbers and interim intervals, using the geometric mean to capture the geometric decrease of the level ks. Nagel (1995) rejects the hypothesis that choices are explained by applying a finite number of steps of the iterated elimination of weakly dominated choices starting at 100.

The model suggests heterogeneity due to subjects' different anticipation or beliefs of what others will do. Level-0 reasoning might be understood as system I behavior, i.e. intuitive response without considering the strategic environment. Higher level ks could be related to deliberate reasoning (system II; Kahnemann, 2011), considering the consequences of others' choices. The simplification in this model is that higher level-k players assume that all others are just one level lower than oneself. Most importantly, the level-k model fills a modeling gap for behavior placed between non-strategic, irrational or random behavior and an equilibrium strategy. It is based on a cognitive reasoning procedure that does not require consistency of beliefs,[26] therefore it is a non-equilibrium model. Yet, in the limit, typically an equilibrium is reached. Most other boundedly rational models we discussed in Section 2.3 do not specifically model heterogeneity, according to our definition spelled out in the introduction, as they assume the same parameterization for all subjects which are distinguished by different random draws from a generated probability distribution.

Georganas et al. (2015) show that level-k distributions across players are stable over many games, but one cannot predict this stability for single subjects across different types of games. However, when presenting variations within a class of games, then the researcher can classify many subjects through one level-k choice, e.g. games are distinguished only through parameter changes such as the b-parameter. (See e.g. Coricelli and Nagel, 2009, doing a brain imaging study with subjects playing standard Beauty Contest games; or Costa-Gomes et al., 2001, where the same subject plays many different 3×3 normal form games with rationalizable equilibria.) Additional players can be classified according to the L-k-rule that most closely corresponds to their choices.

Variations and Extensions of the Basic Level-k Models

Stahl and Wilson (1994, 1995) specify players who best respond to a probability distribution of lower level players in addition to Nagel's level-k types. Camerer et al. (2004) suggest a one-parameter cognitive hierarchy model with types at each level following a Poisson distribution of lower levels, with an estimated parameter of 1.5

[26]Consistent beliefs about strategies means that a player's subjective probability distribution about opponents' play corresponds to the objective probability distribution (see Kneeland, 2015).

being reasonable for a wide class of experiments. See also the surveys by Camerer (2003) and Crawford et al. (2013). Chong et al. (2016) augment the level-k model by stereotype biases, allowing intermediate cases between players with correct beliefs about others' distribution of reasoning levels and a maximum bias with all opponents adopting the most frequently occurring lower level of reasoning.

The following models make important contributions to aspects not covered by the simple level-k versions. Goeree et al. (2016) note that there is a formal connection between level-k and what they call subjective heterogeneous quantal response equilibrium (SQRE). SQRE means that players have non-rational, possibly heterogeneous beliefs about the choices of other players that are inconsistent with their actual choices. A level-k type believes that opponents' choices are concentrated on type $k-1$. While earlier papers like Stahl and Wilson (1995) take the parameter that governs rationality (λ) to be common to all players, SQRE represents the version with heterogeneous skill parameters, so that a higher level of reasoning (k) is associated with a higher skill parameter (λ). Goeree and Holt (2004) introduce the "noisy introspection (NI)" relying on the assumption that choices made under higher orders of the "I think that you think" process are increasingly more difficult to predict. Level-k thus corresponds to the special case in which the rationality parameters for a level-k player take the form $\lambda_0 = \cdots = \lambda_{k-1}$ and $\lambda_k = \cdots = \lambda_\infty = 0$.

A few recent papers have responded to the fact that level-k vary across settings, for instance when beliefs about opponents' sophistications are varied (see, e.g., Agranov et al., 2012; Alaoui and Penta, 2016a; Georganas et al., 2015). The distinction between a player's cognitive bound and his beliefs has been introduced, for instance, in the models of Strzalecki (2014), Georganas et al. (2015), and Alaoui and Penta (2016a). The basic idea in these models is that players with a higher ability of reasoning will react differently in a subject pool with equal minded subjects than with those of lower reasoning types. Players with lower L-ks will not adjust their levels upwards in more sophisticated subject pools. The question of identifying whether given observed behavior is due to "rationality" or "cognitive" bounds is studied by Friedenberg et al. (2015) and by Alaoui and Penta (2017). The meanings of the two bounds are slightly different in the two papers. Friedenberg et al. (2015) distinguish between "bounded reasoning about rationality" and "bounded reasoning about cognition." A subject can be classified as cognitive if she uses some "method" or "theory" as to how to play the game. Being cognitive does not imply rationality, as the subject may use some boundedly rational heuristic. On the other hand, a rational subject is obviously cognitive, as rationality is one possible "method". Friedenberg et al. exploit Kneeland's (2015) ring games to disentangle the two bounds. Alaoui and Penta (2017) instead develop so-called "tutorial" and "replacement" methods, which can be applied to general games, with no special restriction on the payoff structure. Their methods serve to disentangle a subject's cognitive bound, meaning his or her understanding of a situation, from his beliefs over the opponent's own understanding.

The model of Endogenous Depth of Reasoning (EDR) in Alaoui and Penta (2016a) also takes into account that players' cognitive bound may vary systematically with the payoff structure: increasing the stakes of the game induces players to perform more rounds of reasoning. Alaoui and Penta (2016a) provide an experimental test of the EDR model. The EDR model is based on the axiomatic approach developed in Alaoui and Penta (2016b), which provides foundations to endogenizing players' depth of reasoning through a cost–benefit analysis. Cognitive costs also play a role in the rational inattention literature (Sims, 2003), in which agents pay a cost of precisely observing certain variables. In the vein of that literature, one can say a level-k player is (ir)rationally inattentive to higher level players.

A Critical Comment on Level 0

One main criticism against the level-k model concerns the apparent arbitrary or free level-0 specification, e.g. a focal point, naïve behavior or anchor, depending on the situation.[27] Constructing such a variable might be comparable to the identification of reference points in prospect theory (Kahnemann and Tversky, 1979) or the question of how focal points are found (Schelling, 1960) with a need of empirical validation. Another similar concept is salience which has recently attained a lot of interest in the individual decision making literature (Bordalo et al., 2012). "Salience refers to the phenomenon that when one's attention is differentially directed to one portion of the environment rather than to others, the information contained in that portion will receive disproportionate weighting in subsequent judgments" (Taylor and Thompson, 1982). Also learning models need to specify initial conditions of first-period behavior, which are typically exogenous (see, e.g., Roth and Erev, 1995). This question of the starting point of the reasoning process can thus be seen as a bridge to individual decision-making. Do subjects in a game focus first on the situation or the rules as if each one of them was playing in isolation or against nature? Decision theory has produced a tool case of heuristics which is still, to some extent, unexploited by experimenters on interactive decision making. Most applications of the level-k model specify level 0 as an aggregation of uniformly random players' behavior. However, we will also show that other specifications are used in the experiments, discussed below. For some situations, researchers have also suggested multiple reference points to explain the data.

Level-k as a Descriptive and Predictive Model

The level-k model is, first of all, a descriptive model of behavior which provides a classification of different reasoning types and its distribution, given the experimental data. Ex ante one might very well know which are the modal choices, when level 0 can be clearly specified. The most important contribution of this model comes from the large number of empirical observations that most subjects engage in no more than

[27] Brandenburger et al. (2017) write: "A central feature of their identification strategy is that they (authors on level-k) impose auxiliary assumptions about beliefs. This comes in the form of assumptions about the behavior of Level-0 types, which pins down the beliefs of Level-1 types. With this, their notion of iterative reasoning can be conceptually distinct (in subtle ways) from rationality and mth-order belief of rationality (i.e., even in simultaneous move games)."

3 levels. Frequency distributions over these choices are not predicted by the model. Therefore, it should not be possible to predict the average choice, for example. There are, however, ways to have some better indication of future behavior: Obviously, replications of the same underlying experiment should produce similar results regarding the level-k distributions. Thus, knowing the sophistication of the subject pool gives some hints about how behavior may emerge. Most importantly, the goodness of equilibrium as a predictor depends on how many levels are needed to reach equilibrium: in case few levels of reasoning are needed, average behavior can be expected to be close to equilibrium; conversely, if a high number of levels are needed, behavior (above all in early rounds) can be expected to be far from equilibrium. Together with learning rules and the strategic environment (as strategic substitutes or strategic complements), one can indicate whether behavior converges to the equilibrium in the short, medium or long run.

Level-k as a Prescriptive Model

"The problem of prescriptive analysis can be stated as follows: How should a decision analyst advise a client, friend, or himself/herself on how to make a decision?" (Eppel et al., 1992, p. 281). In an interactive decision making problem, one has to consider whether one just consults one player, like "my client" or an entire group. We restrict ourselves here to the one-advisee case. Probably a decent rule-of-thumb advice that can be given to agents is telling them to choose according to level 1 to 2, identifying of course the reference point first. "Don't think too deeply what others are thinking" might be a good advice for real life that has already been formulated by Seneca: "Nothing is more hateful to wisdom than too much cleverness (cunning)." The advisor needs to know the opponents' sophistication or whether players have already interacted for a longer time.[28]

Level-k Modeling in Micro- and Macroeconomic Theory

The stability of limited reasoning (level-k being between 0 and 3), empirically validated through many different experiments, has motivated further work in the micro- and recently also in the macroeconomic theory literature. In applied microeconomics, Crawford et al. (2009), for example, introduce level-k reasoning to the design of optimal auctions. Crawford (2013) use level-k to propose efficient bargaining mechanisms. Kets and Sandroni (2017) extend the level-k model to provide a good model description about how a player's identity affects her reasoning, and use this to study the optimal composition of teams.

In the macroeconomic literature, García-Schmidt and Woodford (2015) introduce a continuous level-k model (which could represent an average of several discrete level-k types) into a standard New-Keynesian model in which households optimize over an infinite horizon. With the introduction of bounded rationality they can justify

[28] Another mention of those thoughts can also be found in E.A. Poe's (1958) "The purloined letter", which distinguishes the mathematician from the poet, and the combination of both, which is ultimately the best reasoner as based on logical reasoning and enough intuition for human behavior.

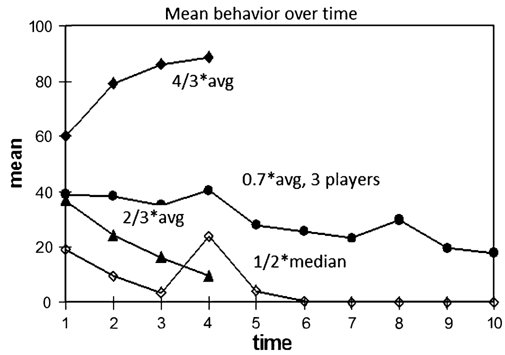

FIGURE 8

Beauty Contest game: Mean behavior over time. Data source: Nagel (1995), Duffy and Nagel (1997) with median; Ho et al. (1998) with 3 players.

the missing increase in inflation after the financial crisis, 2007. A subsequent application of level-k in macroeconomics by Farhi and Werning (2017) introduces level-k into an overlapping generation model with occasionally binding liquidity constraints. Their results differ from García-Schmidt and Woodford (2015), because an overlapping generation model represents a finite horizon. They show that if monetary policy commits to future interest-rate policy, it has particularly weak effects when level-k is combined with incomplete markets. Conversely, under rational expectations, monetary policy would have a rather strong effect. Such papers are now inspiring level-k applications in different macroeconomic contexts. Angeletos and Lian (2016) discuss the relationship between rational expectations equilibrium and "solution concepts that mimic Tâtonnement or Cobweb dynamics, Level-k Thinking, Reflective Equilibrium, and certain kinds of cognitive discounting." They also discuss relaxations of higher-order uncertainty.

4.1.3 Behavior over Time in the Basic BC Game

Nagel (1995) also studies behavior over time, repeating the same game for four periods with the same subjects. After each period, information about all choices, the average and the target number, $2/3 \cdot$ average, and the winning number, are written on the blackboard. Behavior under $b = 1/2$ and $b = 2/3$ converges, albeit at different rates, to zero but stays considerably above zero. If $b > 1$ and $y^i \in [0, 100]$, there are two equilibria: one stable one (100) and one unstable one (0). Nagel shows that if $b = 4/3$, behavior converges to 100, the stable equilibrium. Fig. 8 shows mean behavior over time, slowly decreasing, for different parameter values, group size, and order statistics, i.e. mean or median (Duffy and Nagel, 1997). When the influence of a single player is high ($n = 3$), the mean decreases more slowly than when it is low, as it is the case in the median game.

The level-k model together with a simple adjustment model (called directional learning; Selten and Stoecker, 1986) explain this slow convergence. The idea is that a player in period t, who iterated too many (few) levels from a reference point (say $\frac{2}{3} \times$ average in $t-1$), as compared to the target t, will iterate less (more) in the next period, $t+1$. As a result, even if every subject understands that the target should be zero, there is self-fulfilling slow convergence. Using Nagel's (1995) data, Stahl (1996) combines her level-k model with a "law of effect" learning model: Agents start with a propensity to use a certain level of reasoning, and over time the players learn how the available rules perform and switch to better performing rules. He rejects Bayesian-rule learning in favor of this level-k model.

Fig. 9 shows individual behavior from period t to period $t+1$ for the parameters $b = 1/2$ and $2/3$. Each dot presents a players choice in two consecutive periods. Behavior over time is "flocking" (Raafat et al., 2009), i.e. the variance across players shrinks over time. Flocking occurs faster for $b = 1/2$ than for $b = 2/3$.

4.1.4 Heterogeneity of Subject Pools in Lab- and Field-BC Games

Bosch-Domènech et al. (2002) report on one period BC-games with different populations, including newspaper experiments, in which newspaper readers of the *Financial Times*, *Expansion*, and *Spektrum der Wissenschaft* were invited to participate in a Beauty Contest game similar to Nagel (1995) with $b = 2/3$. Fig. 7 shows the first-period laboratory data (Nagel, 1995), classroom and take home experiments of economics undergraduate in their second year, economists in various conferences, in an online newsgroup experiment and the newspaper contests in Bosch-Domènech et al. (2002).

Behavior is heterogeneous both across different subject pools and within every population. However, there are clear regularities. While undergraduate students hardly ever choose the equilibrium zero, more than 20% of game theorists choose it. On the other hand, outliers selecting 100 or numbers close to 67 are quite similar across populations. According to written comments in many of the BC studies of those subjects who choose 100, they express their frustration of not knowing what a well informed choice could be. This way, they also attempt to increase the average so that low rationality players can win. A choice of 67 can be rationalized since it comes from the game-theoretic property of deleting all (weakly) dominant strategies greater than $100 \cdot (2/3)$. In all datasets, there are clear spikes around 50, 33, and 22, albeit with different frequencies according to the level-k model. Since newspaper readers can reflect upon the game for several weeks, some participants involve parties or newsgroups to participate in the game, to finally submit more informed choices. Indeed those participants had a much higher than average chance to select the winning number. The newsgroup data (last row on the left in Fig. 7) represent a small study, executed by a participant of a newspaper experiment. His choice was less than one integer away from the winning number of our experiment, as he chooses the winning number of his data set.

Bosch et al. (2010) apply a mixture model, which means that the model contains several distributions, to explain behavior in Bosch et al. (2002). They find that four distributions, a uniform distribution, accounting for random behavior, beta distributions for level 1, 2, and infinity (choice zero) explain the actual data best. While

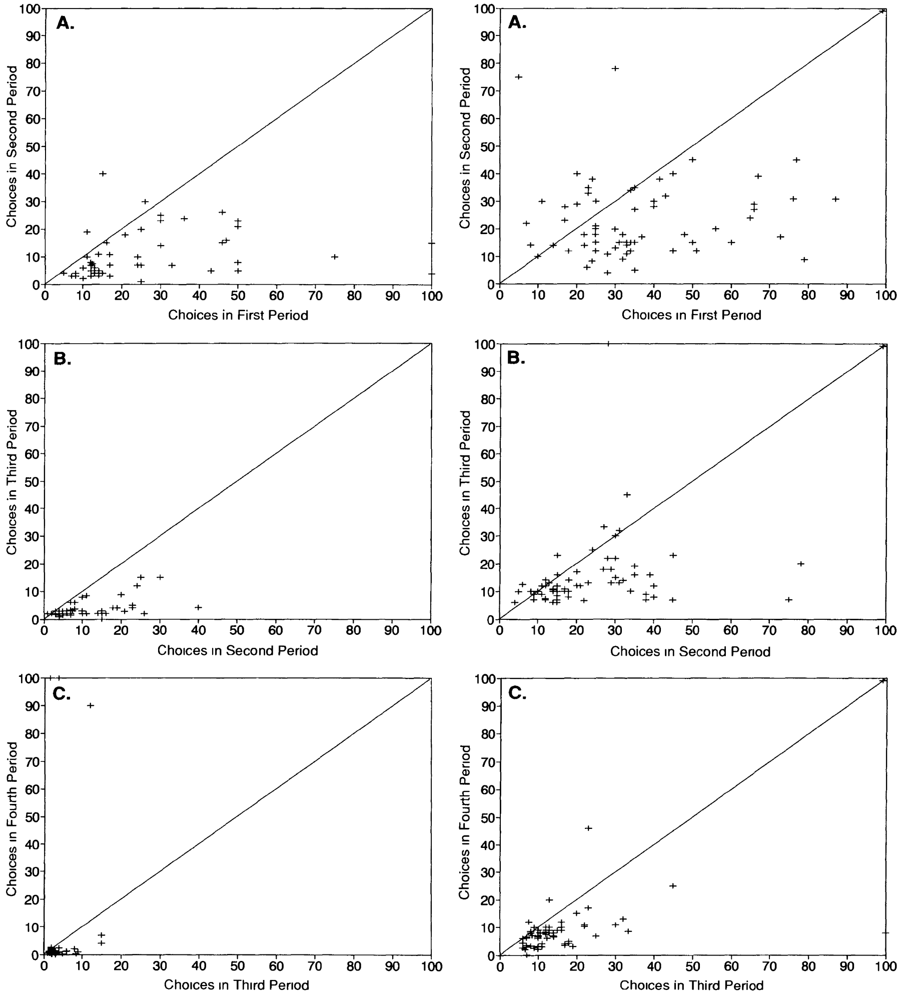

FIGURE 9

Observations over time from periods 1–4, transition between two subsequent periods; left panel: $b = \frac{1}{2}$; right panel: $b = \frac{2}{3}$. Source: Nagel (1995).

all mixtures across different populations have the same means, the frequencies and variances are clearly distinguished across populations. The beta distribution suggests that players do not only have degenerate beliefs and resulting choices, but also tremble. This can be in their beliefs, e.g. they miscalculate or make a shortcut of, e.g., $50 \cdot 2/3$, being 30, or they can tremble in their action, e.g. reaching 22.22 as their second order outcome but choose 25. There is direct evidence for such behavioral results in subjects' comments which the authors requested and subsequently analyzed (see Section 5.5 for more details about this methodology).

4.1.5 Two Person Games

Two Person BC-Games vs $n > 2$-Person BC Game

A natural question is whether players, choosing positive non-dominated choices, are rational and think that opponents are (non)rational or whether players do not understand the logic of the game. Grosskopf and Nagel (2008), Chou et al. (2009), and Nagel et al. (2017) implement several two-person guessing games, in which the target is again $\frac{2}{3}\times$ average. With a tournament payoff, zero becomes a weakly dominant equilibrium strategy, since the lower number of the tuple always wins. However, perhaps surprisingly, Grosskopf and Nagel find that most subjects do not choose zero. Even game theorists playing the game among themselves do not always choose the equilibrium. Chou et al. (2009) repeat the two-person guessing game changing the instructions in simple ways. They demonstrate that behavior is closer to the game theoretical prediction, of zero, if the game is presented in a less abstract way or if hints are given to the subjects, as using the isomorphic game "the lower number always wins". Thus, they conclude that deviations from the game-theoretic solution represent a lack of understanding of the game-form and thus speak of *game-form recognition*. Finding the dominant strategy in this game is cognitively too demanding.

Nagel et al. (2017) resolve the thinking puzzle in such a simple game. With tournament payoffs subjects do not understand that it is sufficient to be closer to the target than the opponent, rather than being as close as possible to the target. The reason is that the best reply to the midpoint (average) of two numbers multiplied with $2/3$ is always pointing to the lower choice. In the "hitting the target" game two players must be paid according to the distance of their choice to the target. As a result, theoretically the iterated elimination of *strictly* dominated strategies leads to the equilibrium, albeit with different elimination steps than if $n > 2$, given the large influence of a single player (that is 100, 50, 25… etc. when $b = 2/3$ and $n = 2$, instead of 100, 66.66, 44.44…). Therefore, zero is not a dominant choice. Nagel et al. (2017) show that the distributions of two persons, $n > 2$ persons, fixed payoff, or distant payoffs all show the same behavioral patterns as observed in the Bosch et al. (2002) study. In addition to that, the professionals also choose in all versions the $n > 2$ patterns, the level-k reasoning with 50 as a focal point, albeit with more choices closer or at zero. The first two boxplots in Fig. 11 show the distributions of choices for $n = 2$ for fixed payoffs and quadratic payoffs with no significant difference. Economics professors (prof) also choose positive numbers (third boxplot).

A rather new method is to let each subject play against oneself in the hope to induce equilibrium choices or less random behavior. In Bosch-Rosa et al. (2016) on the "one-person Beauty Contest", each subject has to choose two numbers between 0 and 100 and the subject is paid for every chosen number according to the distance to $2/3$ of the average of the two chosen numbers (cf. our two-person distance treatment). An unusually high share of subjects (more than 50% of about 350 subjects) find the Nash equilibrium (i.e. the payoff-maximizing answer). Fragiadakis et al. (2017) play two-person Beauty Contest games as in Costa-Gomes and Crawford (2006): Subjects need to be as close as possible to the target, $b \cdot$ *choice of the other player*, where the interval and b can be different for the two players. In some treatments, subjects need to recall (after some other tasks) their choices or play against their own previous play trials 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 non-identifiable patterns. The authors interpret this finding in the following way: level-k reasoning is easier to remember and thus best reply more likely.

11–20 Game

In the original BC-game, the target is quite complicated to understand. Instead, consider the so-called 11–20 ("11 to 20") game, which can be argued to be closer to a real life situation. For example, a tourist may wonder at what time she should leave the beach at the end of the weekend, if she just needs to be on the road a bit earlier than the other drivers, yet wanting to stay as long as possible. Alaoui and Penta (2016a, 2016b) modify the 11–20 game by Arad and Rubinstein (2012a). Two players have to choose a number between 11 and 20, each being paid a dollar amount equal to her choice. Additionally, a player receives a bonus of 20 if her choice is by one lower than the other player's decision. Ties are solved by a random draw. The equilibrium choice 11 is reached by iterated elimination of strictly dominated strategies (20, 19, 8 . . . 11). Level-k behavior is observed with most choices corresponding between L0 and L3 (17–20). The modal choices are either 18 (L2) or 19 (L1). Arad and Rubinstein (2012a, 2012b) observe this kind of behavior, although the equilibrium is in mixed strategies with 15 being the highest probability to be chosen and 20 the lowest. Thus, level-k is a better descriptive model than the theoretical mixing model or the equilibrium 11. An increase of the bonus produces indeed an increase of reasoning, as predicted by Alaoui and Penta who incorporate payoff concerns into the simple level-k model. Because of the ease of understanding this game and at the same time showing important features of level-k, it should be used as a (better) alternative to the basic BC game, especially in populations with known lower cognitive abilities. In such subject pools behavior in BC games are typically concentrated between 35 and 50.

Goeree et al. (2014) consider a variant of the one-shot play of Arad and Rubinstein's (2012a) 11–20 game, where the numbers are arranged on a line with 20 always being the rightmost number (thus being the choice of the level-0 player) and people win if they pick the number to the left of their opponent's choice. They introduce several treatments with different ordering of numbers. For example, if the numbers are ordered in a declining pattern: 19, 18, 17, . . . , 11, 20 then a level-1 player would pick 11. However, the modal choice in this example is 19. The three authors show that the NI-k model helps resolve this puzzle and fits the data better than the level-k model. It turns out that 20 and 19, wherever the later is positioned are the model choices. This can also just mean that they do not understand the rules of the game, but instead those who play 19 order the strategy space according to the natural ascending order.

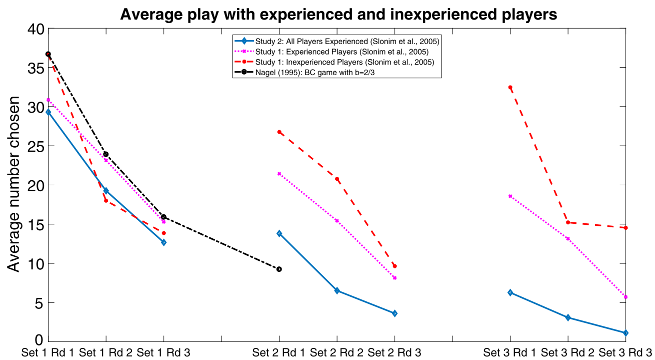

FIGURE 10

Numbers chosen in "guess $\frac{2}{3}\cdot$ average" Beauty Contest game. From Slonim (2005): study 1: new, inexperienced players enter in the course of the game (dashed lines) and experienced players (dotted lines); study 2: all players experienced (solid lines); we added the four periods from Nagel (1995) (dotted–dashed lines). Modified from Slonim (2005).

4.1.6 Mixing Experienced with Inexperienced Subjects

Making informed choices is typically evoked through experience. Slonim (2005) demonstrates that limited cognitive reasoning survives and is optimal when experienced players know that inexperienced ones regularly enter a game. He runs a nine period Beauty Contest game in which only one player stays until the end, while after three consecutive rounds all other subjects are replaced by new ones. Fig. 10 shows that the behavior of newly entering players (the first observation of all line-segments dashed lines) are obviously very similar while informed players (first observations of dotted lines) learn to play a best response to first-period behavior of the new entrants. Such kind of best reply choices are most likely observed in any experiment, in which experienced and inexperienced subjects interact. That is why experimenters typically exclude graduate students and colleagues from their experiments and try to track whether subjects participated in similar games before. In real (non-laboratory) markets, such interactions of informed or experienced vs. uninformed or unexperienced agents are certainly typical which might prevent equilibration. All this can be very well controlled in laboratory experiments. In Section 5, other approaches to improve decision-making such as team reasoning are discussed.

Alaoui and Penta (2016a, 2016b) mix students from more quantitative faculties with humanity students in one of their heterogeneous subject treatments on 11 to 20 game, discussed above. They find that the former group decreases their level of reasoning in the 11 to 20 games against humanity students, while humanity students do not react to such a more sophisticated subject pool. The general finding is that more sophisticated players adjust their level-k downwards when predicting less sophisticated players, and thus heterogeneity emerges within the same subject differentiating between various opponent pools. Less sophisticated players do not respond to more sophisticated players. This indicates that the cognitive bounds of a player seem to be binding as modeled in Alaoui and Penta (2017).

4.1.7 Strategic Substitutes vs. Strategic Complements

Camerer and Fehr (2006), Fehr and Tyran (2008), and Hommes (2013) note that if choices are substitutes, rational agents need to behave in the opposite way to other agents, while if choices are complements, rational agents have an incentive to imitate their opponents. Thus, one can expect that under strategic substitutability outcomes are closer to the rational solution, as less rational behavior is mitigated by rational players, while outcomes are further away from the rational benchmark under strategic complementarities, as less rational behavior is reinforced by rational players.

Sutan and Willinger (2009) study the effects in two different one-shot BC-games with the same interior solutions (discussed also in more detail below), one with strategic substitutes and the other with strategic complements. They find that subjects in the game with strategic substitutes choose the equilibrium much more frequently than subjects in the game with strategic complements. Heemeijer et al. (2009) and Bao et al. (2012) add small or large shocks exogenous shocks to Beauty Contest games and indeed find striking differences in aggregate market behavior under strategic complements, where prices oscillate persistently around fundamentals, and strategic substitutes, where markets quickly settle down to the rational equilibrium. Along the chapter we will give other examples (see also Arifovic and Duffy, 2018).

Group Size

Hanak et al. (2016) investigate how this "strategic environment effect" interacts with group size, conducting each of the two Beauty Contest games with $n = 2, 3, 4, 5, 6$ subjects, larger groups with $n = 8, 16$, and also a situation in which subjects do not know n. They find that as long as groups consist of 5 individuals or more, equilibrium play is more frequent under strategic substitutes than complementaries, corresponding to the earlier result of Sutan and Willinger (2009). Since the "strategic environment effect" is not predicted by the standard level-k, a more sophisticated cognitive hierarchy model is needed to explain the different patterns across games. Alternatively, a level-k (or cognitive hierarchy model) with trembles drawn from a logit distribution can explain the observed results. Arifovic and Duffy (2018) discuss other effects of group size.

4.2 DE-FRAMING THE RULES TO (DE-)ANCHOR BELIEFS

In most studies we cited above, subjects' payoffs only depend on the behavior of all players (typically including oneself), abstracting from many additional features that can influence the behavior and payoffs in more realistic settings. Our above discussion shows that it is very difficult to predict opponents' actions in the first period. Furthermore, the intuitive midpoint together with few reasoning steps produce result in numbers far from equilibrium. However, the equilibrium has ideal properties such

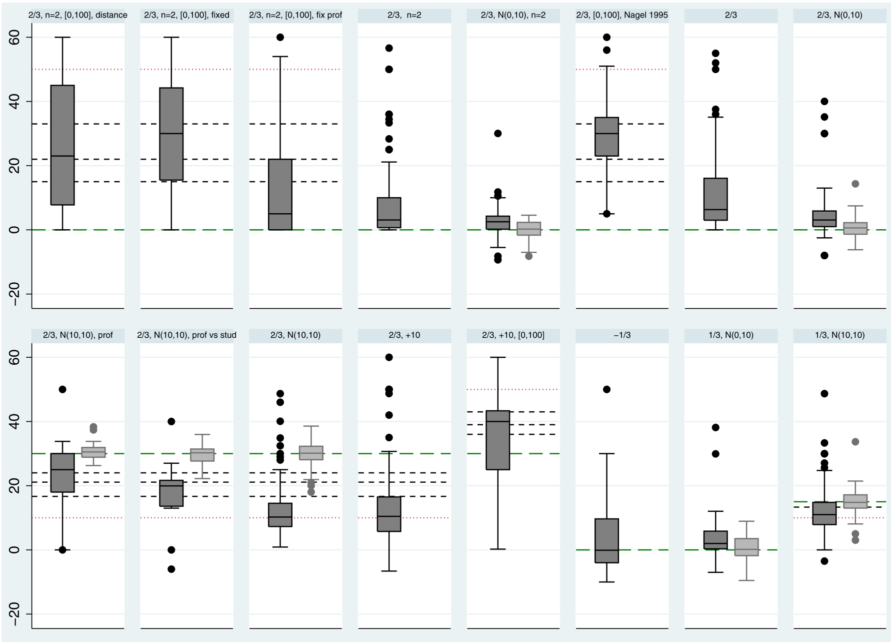

FIGURE 11

Boxplots of different BC-games varying b, $c = 10$, and $N(.,.)$ of ϵ^i, the interval $[0, 100]$ (if not indicated, it is the real line); subject population (professionals = prof. vs. students = stud.); fix = fixed payoff; dist = distance payoff; long-dashed lines represent the (average) equilibrium, and the gray boxplots the equilibrium with noise; dotted lines = L0, short-dashed lines = L1–L3. Data from Benhabib et al. (2018a, 2018b) and Nagel et al. (2017).

as uniqueness and Pareto optimality (in case of a distance function) and equitable payoffs. Thus, the plausible interpretation of deviations from equilibrium is cognitive limitations.

In the remainder of this section, we discuss variations of parameters mentioned in Section 3, to shift behavior in systematic ways.

Fig. 11 provides an overview of some parameter changes we introduce in the following subsections, together with the resulting behavior in boxplots, the equilibrium, and, when adequate, also level-k reasoning. The treatments are shown with parameters $b = -2/3$, $2/3$, and $1/3$; added constants and/or idiosyncratic shocks. The theoretical equilibrium brings about dispersion only with an inclusion of an idiosyncratic noise term, indicated with $N(.,.)$ in the graph. We show that dispersion exists even for subjects that ought to be highly sophisticated, such as economics professors' choices, who were typically invited to participate before a seminar on the topic (prof. = either playing among themselves, or against an existing student distribution, unknown to them (see prof. vs. stud.)).

4.2.1 Interior Equilibrium

Corner solutions have the special property that out of equilibrium behavior can only be in one direction. For example, in a standard public good game there can only be over- but no undercontribution compared to the equilibrium outcome of contributing nothing. In the Beauty Contest games it can also be interesting to move the rational solution into the interior of the strategy space, which we discuss in the following.

An Open Choice Set – Deframing the Rules to De-Anchor Beliefs

Until recently, in most Beauty Contest games the choice sets always constituted a bounded interval, only including positive numbers. Thus, the equilibrium can be attained through an iterated elimination of dominated strategies. Benhabib et al. (2018a, 2018b) allow choices from all real numbers. Those kinds of open sets can be found in reality, e.g. when a person can be a seller or a buyer at the same time and thus prices can be negative or positive. Allowing any number in the choice set of the Beauty Contest game introduces much lower guesses than in sets bounded between 0 and 100 (see Fig. 11, $2/3$, no boundaries vs $2/3$, $[0, 100]$; boxplots 6 and 7, first row). The reason is that in the bounded set, the midpoint serves as a natural anchor and is far away from a corner solution. Without those boundaries, zero becomes a natural focal point, although almost all subjects choose positive numbers when $b > 0$. The median choice decreases from about 33 (Nagel, 1995) to 15 in the open set. However, the focal point and thus level 0 of level-k reasoning becomes zero. This alteration eliminates higher-order beliefs different from zero. Yet, in the experiments not only choices of zero are observed. The newspaper comments of those subjects who find the equilibrium indicate that 80% of the observed choices are between 0 and 10, because they do not believe that others will choose zero. This can explain the choices also in this new setup which is basically "focal or reference point reasoning" with noise. Yet, the noise is only towards higher positive numbers and not negative choices.

Benhabib et al. (2018a, 2018b) introduce b-values that are negative. This provides a change from games with the property of positive feedback to negative feedback, without manipulating other properties of the game. Fig. 11 shows that negative b-values ($-1/3$; boxplot 5, second row of Fig. 11) result in aggregate (average or median) behavior that is much closer to zero than when $b > 0$. Level-k reasoning can easily explain this effect: if I think others choose negative numbers, then my best response will be positive, but if all subjects act in the same way, I should choose a negative number etc. However, choices resulting from higher-order reasoning cannot so easily be distinguished from choices resulting from lower reasoning, as a negative choice can result from L2 or L0. Since more than half of the subjects choose negative numbers, we interpret this as a large share of naive choice, being close to, say, $-2/3$ average, suggesting a negative outcome. Other measures as reaction time or even brain scans can provide more conclusions about the matter, which we discuss in the section on elicitation methods.

Calibration of Coefficients

While the above section discusses the sign of the coefficients, their magnitude matters as well. An important finding that has repeatedly been documented in the experimental literature is that if the b-coefficient is sufficiently close to 1 but still exhibits a unique equilibrium, behavior does not dynamically converge to the rational solution even after many repetitions. Several explanations can be provided for that: first, strategic complements are relatively strong for a high b-coefficient, e.g. $b = 0.95$, so that the effect described in detail by Camerer and Fehr (2006), Fehr and Tyran (2008), and Hommes (2013), mentioned in the above paragraph, that rational players have an incentive to mimic irrational players is particularly distinct. Second, for $b \approx 1$ beliefs tend to become (almost) self-fulfilling. Hence, players are not strongly punished for having non-equilibrium beliefs and those beliefs can be sustained more easily. (See Hommes, 2013 for a more detailed discussion about self-fulfilling beliefs.) Third, even small mistakes or misperceptions of b may suffice to introduce a unit root into the process, under which outcomes would not converge. (See Mauersberger, 2016 for a more detailed discussion.)

Sonnemans and Tuinstra (2010) conduct simple Beauty Contest game experiments, setting $b = 0.67$ in one treatment and $b = 0.95$ in another treatment. Independently of whether they provide information about the b-parameters to their subjects or not, they find fast coordination among the cross-section of subjects for $b = 0.95$, but no or only slow convergence to the rational solution in those groups. Conversely, groups with $b = 0.67$ converge fast to the rational solution, while there is much heterogeneity in the initial periods. Non-convergence in most groups for $b = 0.95$ has also been found by Heemeijer et al. (2009). Bao and Hommes (2017), conducting a learning-to-forecast experiment with a framing related to housing markets, have three treatments with b-values of 0.95, 0.86 and 0.71. For $b = 0.95$, they find explosive behavior, for $b = 0.86$ persistent fluctuations and for $b = 0.71$ fast convergence to the rational solution.

4.2.2 Public and Private Information – Anchoring Beliefs

Beauty Contest Games with a Known Constant

A simple way to move the efficient equilibrium in a Beauty Contest game into the interior, was first introduced by Güth et al. (2002). Subjects are told to be close to $y^i = c + b\hat{E}^i \frac{1}{N}\sum_{j=1}^{N} y^j$ with a bounded interval 0 to 100. They find that the commonly known constant $c = 50$ is chosen by many participants in the first round, but also report some subjects choosing close to $c \times 2/3$ or $c \times 2/3 + c$. The equilibrium in this game is 60. Choices therefore converge fast to 60 in repeated rounds. However, choosing c as the midpoint, does not change the focal point of the basic BC game.

Therefore, Benhabib et al. (2018b) provide two new treatments with $c = 10$: firstly, only allowing subjects to choose numbers within the closed interval 0 to 100, and secondly, allowing another set of subjects to choose from all real numbers (see also Fig. 11, boxplot second row, 2/3, +10; and 2/3, +10 [0, 100]; and the horizontal short-dashed lines which represent levels 0 to 3; equilibrium 30). Within a closed boundary, 50 remains the focal point and most numbers are between to 30 (about $50 \cdot (2/3)$) and 43 ($50 \cdot (2/3) + 10$), consistent with Güth et al. (2002) findings of the reference point being 50. However, without a boundary, most subjects seem to choose 10 as their level-0 belief; or level 1 (two possibilities: near $16 = 10 \cdot (2/3) + 10$; or alternatively, $6.66 = 10 \cdot (2/3)$; or L2 (near 21), instead 30, the equilibrium). The resulting average behavior is around $c = 10$ and thus most choices being between level 0 and 1. We want to stress here, how sensitive behavior reacts to the specifications of the rules: changing the choice set, from boundary to non-boundary, shifts average behavior from about 35 to 10. Furthermore, the anchor c does not provide instantaneous equilibrium behavior, as it is not near or equal to the equilibrium choice.

Common but Unknown Constant

Shapiro et al. (2014) as well as Baeriswyl and Cornand (2014) test the Morris and Shin (2002) Beauty Contest model with a common unknown additive constant c to the BC game. They find that level-k is a good descriptive model in standard settings when information is symmetrically distributed among players, giving only private information. However, subjects' behavior in the asymmetric case is mainly driven by focal points such as choosing halfway between the public and the private signal, or choosing one of these two signals. Thus, whether the equilibrium is instantaneously learnable depends very much upon whether naive players receive an equilibrium strategy as a reference point or not. Behavior over time remains heterogeneous since off-equilibrium payoffs are rather flat and similar to equilibrium payoffs, which impedes learning towards the equilibrium.

Idiosyncratic Precise Shocks with Zero Mean – Equilibrium Anchors

Benhabib et al. (2018a, 2018b) add to the b-average component an idiosyncratic payoff relevant shock ϵ^i drawn from a normal distribution with zero mean, implementing the abstract model of a general equilibrium with sentiments (Benhabib et al., 2015). In each of the eight periods, subjects know the parameter b, differing in each period, and each receives a privately known idiosyncratic shock. There is no feedback between periods, in order to study behavior without learning through realized payoffs. Since drawn signals could result to be negative or outside a boundary, the experiment implements an unbounded choice set, the real number line. Subjects' payoff is given by a quadratic distance function between their choice, y^i, and the target as given in Eq. (3). The experimental results demonstrate that the signal ϵ^i becomes an individual anchor, being indeed the modal choice with the average choice being close to zero, for all, positive or negative, b-parameters. Thus, level-k reasoning finally breaks down: in equilibrium everybody has to choose her signal, and subjects at least on average do so, such that sunspots can be considered a subtle reference point system. This is because these signals coordinate the behavior of strategic and naive subjects: both make the same kind of choice and earn high payoffs. This signal structure can be called "strategic nudging". Fig. 11 shows some typical patterns for $b = 2/3$ and $b = 1/3$ and an added ϵ^i (see row 1, boxplots 5 and 8, and rows 2, 7, and 8), with a mean close to zero.

Idiosyncratic Noisy Shocks

In the same experiment of the previous paragraph, in a slightly modified setup (also based on Benhabib et al., 2018a) players receive correlated noisy signals about the idiosyncratic payoff-relevant shocks ϵ^i. The private noisy signal (called hint) given to each subject is a convex combination of the private, payoff relevant shock ϵ^i and a common shock z, with both shocks drawn from two normal distributions with zero mean. Neither the realization of the private shocks nor the common shocks are known independently. The parameters are constructed such that in equilibrium a player chooses $1.5 \cdot hint$. Here, on average, subjects choose about $0.5 \cdot hint$, which is closer to its equilibrium than average choices in the original Beauty Contest game with an open choice set. Yet, this new game represents an analytically more sophisticated problem. The reason for higher payoff outcomes is that the majority chooses from three kinds of simple heuristics: the modal group of subjects ignores the hint altogether and chooses zero; some choose (near) the noisy signal, thus treating the signal as if it were precise; or others compute a choice resulting from a hint, expecting that the common component shock z is zero. Since the noisy signals become focal points for subjects but are at the same time known from the perspective of the analyst, they are a suitable feature for predictive behavior. The analytical solutions of games with signals are, however, more sophisticated, especially with noisy signals, than in the original Beauty Contest game.

Order Effects from Different Idiosyncratic Shock Treatments Within the Same Subjects

Benhabib et al. (2018a) have subjects play several b-Beauty Contest games with precise or imprecise signals without information between rounds, and then the b-Beauty Contest games without signals. The resulting means in those b-Beauty Contest games without signals are then close to zero, as if subjects played a game with a public announcement "$\epsilon^i = 0$" for all players.

Therefore, the implementation of signals in Benhabib et al. (2018a, 2018b) can be considered to solve Keynes' level-k "problem" by proposing Keynesian sentiments denoted by ϵ^i, to alleviate the need of subjects to do complicated calculations. They just can use their anchor. The analytical treatment of this game is, however, more sophisticated, especially with noisy signals, than in the original Beauty Contest game. In the next paragraph we show, however, that when these added shocks are not equilibrium choices, then behavior is again distorted from the efficient equilibrium as it has already been the case in the treatments with the commonly known constant c.

Idiosyncratic Shocks with Nonzero Means

Benhabib et al. (2018b) add a treatment in which idiosyncratic signals, ϵ^i, are drawn i.i.d. from $N(c, \sigma^2)$, $c > 0$. The resulting equilibrium is $(b \cdot c/(1 - 2/3)) + \epsilon^i$ (see Benhabib et al., 2018a, 2018b). This means, if e.g. $c = 10$, then one should choose $20 + \epsilon^i$. However, the authors show that subjects choose their signals as an anchor, ignoring the effect of the mean signals on the average; some also just choose $2/3 \cdot \epsilon^i$ or $2/3 \cdot \epsilon^i + c$ which both can be considered as level-1 reasoning as in the case of the treatment $2/3 \cdot avg + 10$. If subjects are Economics professors, average play is higher, but still different from the equilibrium mean. When each professor plays against the student populations they choose lower numbers (and thus lower level ks), and typically one step higher than the student population, as if they played "best reply" against this student population, anticipating that students play around c. Fig. 11 shows these treatments in the second row, $\frac{2}{3}$, $N(10, 10)$, which is isomorphic to the game $\frac{2}{3} + 10$ with $N(0, 10)$, and also the behavior is the same.

Bayona et al. (2016) conduct a laboratory three-person supply schedule experiment with privately known (un)correlated costs. In the Bayesian equilibrium, predictions for positively correlated costs result in steeper supply functions, and thus more collusive behavior and greater market power than do uncorrelated costs. Subjects receive private signals about the (un)correlated costs. They find that most subjects bid as predicted by the equilibrium in the simple, uncorrelated cost treatment, where the constant term of the linear supply schedule chosen by the subjects is around the own signal cost. However, in the more sophisticated, correlated cost environment their bids display large heterogeneity and considerable deviations from equilibrium with the average behavior being close to the uncorrelated case. The reason for these similarities are easy to understand: in the uncorrelated case subjects are asked to reason in a simple strategic environment. In the correlated case, subjects largely ignore the feature of correlation or are unable to incorporate it in their reasoning. As a consequence, best response behavior results in behavior close to the uncorrelated case. Since naive subjects still make some profits, learning is impeded and it is optimal for sophisticated players (who understand the correlation implication) to mimic unsophisticated behavior.

4.2.3 Stochastic Common Shocks

Common Time Varying Shock

Another version of the BC game that has been implemented in the lab is $y_t^i = b\hat{E}^i \frac{1}{N} \sum_{j=1}^{N} y_t^j + \epsilon_t$. In this version ϵ_t is a commonly known realization for all participants. Over time, ϵ_t could for example be determined by a random walk: $\epsilon_t = \epsilon_{t-1} + \xi_t$. In this case, the experimental literature documents that the convergence in a BC game with a common and known constant found by Güth et al. (2002) is severely mitigated.

Lambsdorff et al. (2013) implement a price setting game with the target price of $4/5 \cdot (\text{average} + 5) + \epsilon_t/10$ with 5 being explained to the subjects as the "cost of a raw material" and ϵ_t as a "business indicator" being a randomly selected integer from the interval $[-15, 15]$.[29] Subjects choose the business indicator as a simple heuristic ϵ_t instead of the equilibrium $(20 + \epsilon_t/2)$. This shows again that reference points can wrongly anchor subjects' behavior.

In another Beauty Contest experiment, Giamattei and Lambsdorff (2015) phrase the game as a macroeconomic model of the Keynesian multiplier with a time-varying

[29] In their paper, this variable is called c. We refer to it as ϵ to stress that this variable is stochastic.

component ϵ_t being an exogenous, non-stationary sequence of investments driven by random shocks. Participants have to decide upon their individually optimal consumption in a way that they have to be close to $4/5 \cdot$ income, where $4/5$ is the marginal propensity to consume. Income is calculated as the average consumption over all subjects + the investment ϵ_t. Again, subjects choose the investment component as an anchor. The authors implement two treatments: Firstly, in the so-called Keynes-treatment, subjects are punished both for upward and downward deviations, while, secondly, in the White-treatment they are only punished for upward deviations of consumption, meaning too low savings. This setup reflects the debate of Keynes and White whether only deficits on the current account should be punished as argued by the latter. Keynes instead advocated punishing both surpluses and deficits.

Giamattei and Lambsdorff (2015) find that due to boundedly rational reasoning, asymmetric punishment of deficits leads to permanent under-consumption. To the best of our knowledge Keynes himself did not apply his own Beauty Contest reasoning process to the model of consumption behavior in games with the so-called Keynesian multiplier. Instead, he pondered that consumers can reason their way to equilibrium instantaneously (thus infinity reasoning), given a simple maximization argument. “Net income is what we suppose the ordinary man to reckon his available income to be when he is deciding how much to spend on current consumption” (Keynes, 1936, p. 56). Yet, in his lecture notes Keynes (1973, p. 181) deplored his approach: “I now feel that if I were writing the book again I should (...) have a subsequent chapter showing the difference it makes when short-period expectations are disappointed.”

Giamattei (2017) uses the BC framework with a common time varying shock to test the role of limited reasoning on inflation inertia and disinflation. In his game four price setters have to set inflation rates which are strategic complements with $b = 2/3$. The added random process ϵ_t reflects the part of the inflation rate which can be set by the central bank. The task of the central bank as a fifth player is to reduce inflation. Giamattei (2017) shows that Cold Turkey, a sudden change in ϵ_t, does not help to quickly reduce inflation as proposed by theory. Instead subjects show a large degree of inertia which could be better accounted for by a more gradual approach of lowering ϵ_t.

Guessing a Time Varying Shock

Another (extreme) situation which is independent of other players can be represented by a Beauty Contest equation of the form $y^i = \epsilon^i$ where $c = f(.) = 0$ and ϵ^i being a random element chosen by the computer. In that case, the task becomes an *individual decision-making problem.* In the next paragraph we give one out of possibly many examples.

Khaw et al. (2017) conduct an individual decision-making experiment in which subjects need to estimate the probability of drawing a green ring with replacement out of a box with green and red rings. Subjects see each period a draw and submit their draw-by-draw estimate. This probability can be considered to be a continuous random shock, since it is computer-generated and varies stochastically. They find that participants systematically depart from the optimal Bayesian prediction and can be explained by a model of inattentive adjustment. On average behavior is close to Bayesian reasoning with a high variance.

4.3 MULTIPLE EQUILIBRIA (BC GAME WITH $b = 1$)

Beauty Contest games with $b = 1$ represent the original Keynesian Beauty Contest game. The target is typically to choose what most others select. In this setup, there is no higher order reasoning necessary. If a player believes that others, on average, choose a number x, then her best response is also x. However, as stated in the level-k model, the question is what most players pick, as e.g., the focal point (level 0), or what a naive person would play.

4.3.1 Minimum Effort Game

In Van Huyck et al. (1991), players' payoffs depend on their own choice (1, 2, ..., or 7, which enters as a cost) and the minimum choice of all players (which enters as a benefit). Any choice can form an equilibrium, chosen by all. This can be labeled as a minimum effort game, as the payoff is the greatest if all players choose 7, while it is the least if at least one player chooses one. Heterogeneity in behavior thus induces efficiency losses in this game. For groups of size 14 to 16, there is heterogeneity in the first period as there are many different arguments about optimal play, given the belief of a possible minimum play. After the third round, in all sessions the minimum was 1 with a simple reasoning: best response to a minimum 1 in the previous period is to play 1. This is very different when playing repeatedly in groups of two persons which fostered coordination to the efficient equilibrium a great deal, since 12 out of the 14 groups achieved this equilibrium. If one sees that the other player played a high number, one is inclined to play a high choice as well, with little mismatch within a pair.

Fig. 12 depicts the choices in period 1 for all treatments in Van Huyck et al. (1991) median effort game. In the baseline treatment that combines Gamma and Gammadm, there is one payoff-dominant (7, ..., 7) and one secure equilibrium in the set of strict equilibria (3, ..., 3). In the first period, 70% of the subjects choose neither of the two but instead an action between the payoff-dominant and the secure equilibrium. Over time, behavior converges to the inefficient equilibrium selected by the "historical accident" of the initial median choice. In the second treatment, Omega, in the set of strict equilibria there is only a payoff-dominant equilibrium (7, ..., 7) and no secure equilibrium, which is chosen by 52% of the subjects in the first period. In the third treatment Phi, in the set of strict equilibria there is only a secure equilibrium (4, ..., 4), which is the modal choice played by 41% of all players in period 1. Over time, the equilibrium is selected by the "historical accident" of the initial choice.

Experimenters have investigated which institutional designs can increase efficiency in these basic coordination games. We will present just a few such attempts. Van Huyck et al. (1991) vary the payoff structure so that they can investigate two salient equilibrium selection principles: payoff-dominance, being the equilibrium

TABLE II
DISTRIBUTION OF CHOICES IN PERIOD 1

	Treatment									
	Gamma		Gammadm		Combined (baseline)		Omega		Phi	
Action	Nm.	(Pr.)	Nm.	(Pr.)	Nm.	(Pr.)	Nm.	(Pr.)	Nm.	(Pr.)
7	5	(18)	3	(11)	8	(15)	14	(52)	2	(7.5)
6	3	(11)	1	(4)	4	(7)	1	(4)	3	(11)
5	8	(30)	7	(26)	15	(28)	9	(33)	9	(33)
4	8	(30)	11	(41)	19	(35)	3	(11)	11	(41)
3	3	(11)	5	(18)	8	(15)	0	(0)	2	(7.5)
2	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)
1	0	(0)	0	(0)	0	(0)	0	(0)	0	(0)
Total	27	(100)	27	(100)	54	(100)	27	(100)	27	(100)

Notes. Nm. = number of subjects. Pr. = percent of subjects.

FIGURE 12

Table II in Van Huyck et al. (1991).

that would give the largest payoff, and security, meaning that choosing an action whose smallest payoff is at least as large as the smallest payoff for any other action. Weber et al. (2001) investigate whether coordination could be achieved by randomly appointing a group leader who encourages the group to select large numbers. This limited one-way communication does not improve efficiency, since large (8–10 subjects) groups are not able to coordinate towards the efficient outcome, and small (two subjects) groups attain efficiency independently of whether they had the leader's announcement or not. Camerer (2003) surveys those kinds of coordination games.

4.3.2 Public and Private Signals

Fehr et al. (2017) investigate multiplicity of equilibria in an experiment, where two agents in a repeated interaction of 80 periods need to simultaneously choose an integer from 0 to 100, being rewarded according to a quadratic loss function, penalizing their distance from the other player's number. The extrinsic information was displayed as a randomly-drawn integer in the interval from 0 to 100. Each player in turn receives a signal of this number, being the same as this number with a fixed probability p. The treatments differ according to whether the signal is public or private or both and according to the preciseness of the signal: the baseline (N) is without signals; in treatment P75 (P95) subjects receive a private signal with a precision probability of 75% (prob = 95%); in treatment AC subjects receive a precise private signal but the signal is only revealed to each subject with probability 90%; in treatment C subjects receive a public signal with precision probability of 75%; in treatment CP subjects receive a public and a private signal, both with a precision probability 75%; in treatment CC subjects receive two public signals.

The results, being depicted in Fig. 13, show that public signals are not strictly necessary for sunspot-induced behavior or coordination. Even in the absence of extrinsic

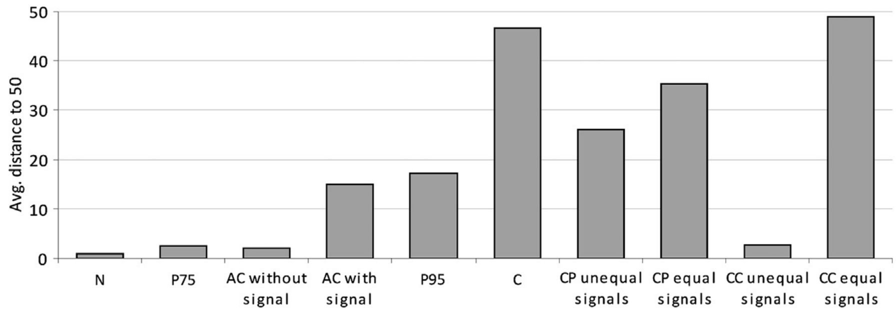

FIGURE 13

Fehr et al. (2017, Fig. 3): Average distance of choices from the secure action (avg. distance from 50).

information (treatment N and the observations of treatment AC without signal), subjects tend to coordinate on choosing 50, the middle of the interval. If the signal is imprecise (P75) or if two signals are difficult to aggregate because they are not highly correlated (CC unequal signals), the focal point of 50 dominates in determining subjects' choice. On the other hand, if the signal is precise (P95), there is one signal (C) or the signals are highly correlated (CC equal signals), there is a clear tendency to follow the action indicated by the signal. If a public signal is combined with a private one (CP), the ability to coordinate behavior on any number is impeded and payoffs are lower on average.

The coordination result resembles one particular finding in the experimental macroeconomics literature (discussed in more detail by Arifovic and Duffy, 2018). In the learning-to-forecast experiment in a New-Keynesian framework by Assenza et al. (2014), one treatment is run with an inflation targeting interest rate rule with coefficient 1, which induces a continuum of RE steady state equilibria. They find coordination on one of these arbitrary equilibria or coordination on unstable inflationary or deflationary spirals.

4.3.3 Global Games

As explained in Section 3, payoff disturbances in coordination games with multiple equilibria can theoretically lead to a unique equilibrium. Heinemann and Noussair (2015) provide a more extensive survey of this literature. Cabrales et al. (2007) test the global game theory based on Carlsson and Van Damme (1993) in a two-person game with two actions for each player and random matching between players each period. In the complete information control treatment, many pairs converge quickly to the Pareto optimal equilibrium while the minority converges to the risk dominant one. In the incomplete information case, each player receives an idiosyncratic noisy discrete signal about the actual payoff of the safe action. Payoffs for the alternative and risky action, is either zero when the other player chooses the safe action or 15 if both players choose the alternative, which is the Pareto-optimal equilibrium. After some periods of miscoordination, almost all pairs coordinate on the global game solution, the risk dominant equilibrium.

For groups of 15 subjects, Heinemann et al. (2004) test the main prediction by global games departing from the speculative currency attack model by Obstfeld (1996) and Morris and Shin (1998). At the beginning of each period, a payoff-relevant random variable Y is drawn from a uniform distribution with known support. Y can be interpreted as the fundamentals of the setup so that higher (lower) values of Y represent worse (better) fundamentals.

There are two treatments: complete information (CI) where Y is known to all 15 subjects and private information (PI), where Y is imperfectly revealed to each of the 15 subjects by idiosyncratic noisy signals. These signals are drawn randomly from a uniform interval $X_i \sim [Y - \epsilon, Y + \epsilon]$ with ϵ being small. Agents then have to choose between two actions, A and B: A, which can be interpreted as "not attacking", is a safe choice yielding a fixed payoff F. The other action, B, which can be interpreted as "attacking", yields a payoff which depends on the total number of subjects choosing B. The critical threshold is given by a monotonically decreasing function $a(Y)$: should less than $a(Y)$ players choose B, all who choose B earn 0, while if more or equal to $a(Y)$ players choose B, the attackers earn a payoff equal to the fundamental Y.

Testing the theory requires that the distribution of Y values is chosen so that there exist values of $Y \leq F$, for which the dominant strategy is choosing A and similarly there exist values of $Y \geq a^{-1}(1)$ so that an individual subject is incentivized to attack by selecting B. Under complete information, there are multiple equilibria, either all players choosing A or all players choosing B, for Y lying in the interval $(F, a^{-1}(1))$. Once information is incomplete, there exists a unique threshold value of the noisy signal above which all subjects should choose B, i.e. to attack, and below which they should not attack and thus choose A.

The main aim of Heinemann et al. is twofold: first, comparing the stability of the complete information game with its multiplicity of equilibria to that of the private information game; second, whether subjects use threshold strategies that are predicted by the global game threshold. The results show that subjects play the threshold strategies in both the private and complete information treatments and estimated thresholds in general lie below the global game predictions X^* or Y^* but are higher than the payoff-dominant strategy of choosing B if $Y > F$. In contrast to Cabrales et al., Heinemann et al. find much coordination over time on a common threshold and therefore little heterogeneity in the complete information treatment, where Y is publicly known and there are in theory multiple equilibria. Conversely, in the incomplete information treatment there is less coordination because of variation of signals between players and therefore more persistent heterogeneity in behavior.

Another paper that tests global game theory is Cornand (2006), who adds two treatments, one where subjects receive a private and a public signal and a second in which subjects receive two noisy public signals. She finds that subjects tend to overreact to the public signal even when they also receive a private signal. However, subjects' behavior tends to be more random if they receive two noisy public signals.

This result implies that if institutions make public announcements, they would be well-advised to coordinate on one single message.

Heinemann et al. (2009) use the full information original (2004) design but additionally, collect data on subjects' degree of risk aversion and their subjective beliefs about the decisions of other subjects. This additional information is required to compare two possible channels of strategic uncertainty in the context of the global games refinement: uncertainty about monetary payoff versus uncertainty about risk attitudes. Their findings are as follows: firstly, adding to a behavioral model either noise to the payoffs or heterogeneity about risk attitudes, performs well in terms of in-sample and out-of-sample performance; secondly, more risk-averse agents are less likely to play the risky choice B; thirdly, players under- (over-) estimate the success probability of coordination when a low (high) number of players choosing B is required to exceed the threshold. In such kind of games of strategic complements players just have to best respond to their subjective beliefs of B-choices of others, which is therefore just level-1 reasoning.

Duffy and Ochs (2012) extend Heinemann et al.'s (2004) design by introducing a dynamic setting so that players repeatedly play the game deciding whether to attack or not, given the history of previous decisions of other players. The result is that even if there is a cost for waiting to attack, they do not find much difference between the thresholds in the dynamic game and in the static game. This is good news in terms of that the simultaneous attack game is a good approximation of the dynamic game.

Szkup and Trevino (2011) also embark on the Heinemann et al. (2004) design to investigate how costly information acquisition impacts on behavior in global games. Here subjects can choose the precision of the private signal received, where more precise signals are costlier. Their results display large heterogeneity, showing that only 30% subjects choose the equilibrium middle level of precision and as opposed to their theory, there is a positive correlation between choosing more precise signals and the frequency of choosing to attack.

4.3.4 Ultimatum Games (Played in Extensive Form)

We described in Section 3 an ultimatum game with simultaneous moves, which is a Beauty Contest game with proposer's demand for himself equals $b = -1$ times the choice of the opponent plus a constant that presents the size of the pie. Typically, this game is played in extensive form, i.e. the proposer sends an offer to the responder, who then reacts only to this stated offer. Experiments like Roth et al. (1991), conducted in several countries, show in two-person games between 50% and 70% of proposers' offers (pie size minus demand) lie between 40% to 50% of a given pie, which is (near) the focal point, an intuitive answer (thus level-0 interpretation), or can also be seen as a best response to the (equal split) norm, expected as the responder's acceptance level (thus level-1). Offers below 30% are typically rejected by the responder. Over time, behavior remains rather stable tilted towards equal splits. However, there are cultural differences across different countries. For example, Japanese

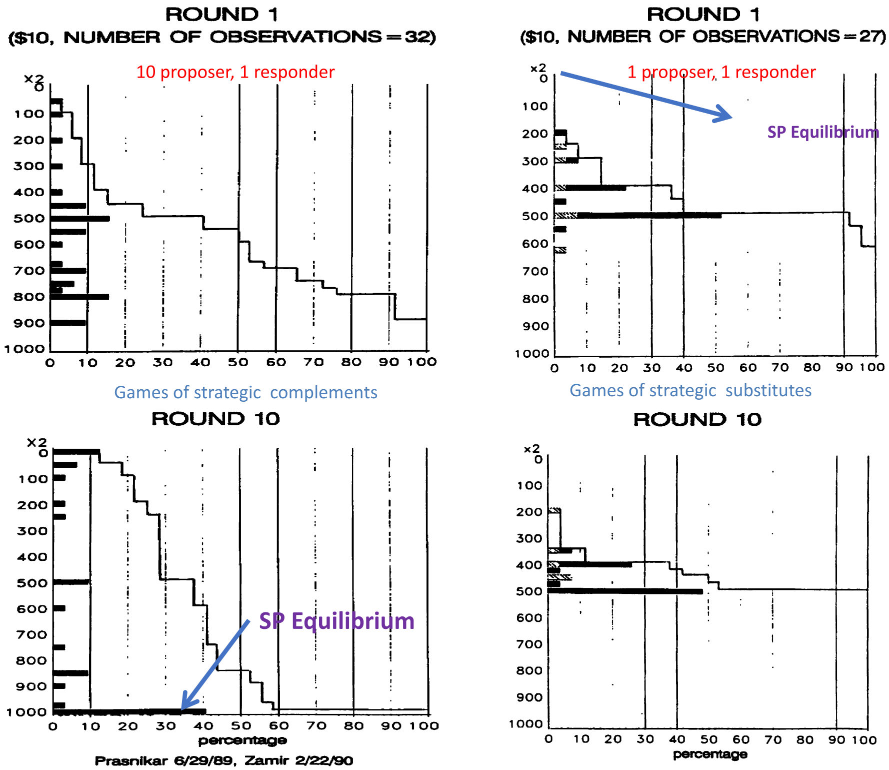

FIGURE 14

Roth et al. (1991): Distribution of accepted (black part of columns) and rejected (crosshatched part of columns) market offers in the United States. The left figure presents 10-proposer markets and the right figure markets with one single proposer.

students choose lower offers which are also accepted, given that the norm is somewhat different as already discussed in Section 2 (bargaining).

Roth et al. (1991) also increase the number of proposers in an ultimatum game and thus turn the game into a Bertrand game with strategic complements with no cultural differences. Behavior converges to the unique subgame perfect equilibrium (40% of offers result in SPE-play), giving almost all of the share to the responder. Analogously, competing responders (Grosskopf, 2003) induce small proposer offers, which are accepted as responders would find themselves unable to punish self-regarding proposers unilaterally. Fig. 14 presents the first and the 10th period distributions of accepted and rejected offers for the 10 proposers (left figures) and one single proposer (right figures). Over time, there is more heterogeneity in the 10 player games than in the one player game, since in the former resulting equilibrium payoffs are rather small, and thus deviations are not punished.

4.4 AUCTION IN THE LABORATORY AND THE FIELD

Here we give some examples of auctions experiments and the interpretation of the data through the level-k model (for an extended discussion see Crawford et al., 2013 and Crawford, 2013).

LUPI

An experimental game conducted in the field that shares the features of the lowest unique bid auction is the LUPI (lowest unique positive integer) game. LUPI is a Swedish lottery game, asking individuals to choose an integer from 1 to K. The winner is the person choosing the lowest integer that nobody else picks. The equilibrium varies depending on the number of players. Östling et al. (2011) calculate the Poisson–Nash equilibrium, a mixed equilibrium, assuming the number of players playing the game every day, is Poisson-distributed. They find that observed play in both laboratory and field data is fairly close to equilibrium play over time. To explain the heterogeneity in player choices, they use a cognitive hierarchy model by Camerer et al. (2004) which assumes a poison distribution by a level-k play over all lower levels.

Auctions with Private Signals

Georganas and Nagel (2011) use a two-player English clock common value auction of a take-over game in the laboratory to test whether agents with a toehold advantage (one owning more shares of the firm than the other) bid more aggressively. The intuition for this theory is the following: imagine bidders want to acquire a company in an auction, in which the value corresponds to the sum of the two private values of the two bidders. One player owns already a larger share (toehold) of the firm than the other. If the player with the larger share wins the bidding, he does not have to pay the full price; if he loses, he receives a lucrative compensation from the winning party. In equilibrium the low toeholder bids close to his signal and the large toeholder has an exploding bid, involving a high level of reasoning to reach the equilibrium bid. Actual larger toeholds increase the winning probability and the payoff of their owners. While these results are in line with theory, they are not nearly as strong as predicted by theory. Bids of higher toeholders are only significantly higher than of lower toeholders, if they own more than 20% of the shares. Behavior can be explained by the level-k model: both bidders bid according to level 1 types, where level 0 is bidding equal to signal value (truthful bidding). This kind of level 0 is equivalent to assuming that the level 0 bidder chooses uniform randomly over the known signal interval. There is also no learning over time. The authors explain this finding by the fact that the payoff function does not penalize deviations from the equilibrium harshly enough.

Crawford and Iriberri (2007) explain behavior in first price private value and first and second price common value with the winner's curse (winning the object, but loosing money), using a level-k reasoning model. They conclude that level-k has much of the same implications as equilibrium auction theory. Furthermore, a level-k model based on types up to L2 can explain the winner's curse in common-value auctions and overbidding in private-value auctions. The intuitive explanation for this is that level-0 types randomly choose a bid or are truthtelling, which can provide an anchor above the equilibrium. Level-1 types anticipate this and give a best response, also being above equilibrium. Crawford and Iriberri also show empirically that a mixture level-k model up to level 2 outperforms the "cursed equilibrium model" by Eyster and Rabin (2005) using three experimental datasets: Kagel and Levin (1986), Avery and Kagel (1997), and Goeree et al. (2002).

4.5 OTHER GAMES WITH DISCRETE STRATEGY SPACES

4.5.1 A Ring Game – Spatial Reasoning

Kneeland (2015) implements a new experiment based on a ring structure which does not require the experimenter to specify level-0 players. In her ring structure with four players, each player chooses from three possible actions. Player 1's payoff may depend on player 2's action, while player 2's payoff may depend on player 3's action; and player 3's payoff depends on player 4's action and player 4's payoff depends on player 1's action. A player with first-order rationality only considers her belief about her neighbor's action, while a player with second-order rationality considers the neighbor's belief about her neighbor's action etc. Since optimal actions differ according to the depth of reasoning, this allows inferring the depth of reasoning of different players: Kneeland (2015) finds that 93% of subjects are at least first-order rational, 71% are at least second-order rational, 44% are at least third-order rational and 22% are at least fourth-order rational. Note, that the author formulates the belief reasoning along the spacial component due to the network structure.[30]

Friedenberg et al. (2015), using the data of Kneeland (2015), find that a large share of subjects that apply reasoning about cognition does not necessarily apply reasoning about rationality. More specifically, approximately 50% of the subjects that are first-order or second-order rational can be considered as applying higher levels of cognition.

Decomposing the (average) behavior into several types made it possible to understand further deviations or (non)-convergence to plausible, efficient or other mathematical solutions. This in turn gave the researchers ideas to construct new designs obtaining different (favorable) outcomes. This will be discussed in the next sections.

4.5.2 Schelling Matching

Benito et al. (2011) let subjects play Schelling's spatial proximity model. There are two types (that ought to represent for example white and black), to each of which N subjects are assigned. The neighborhood of each subject consists of the right-hand

[30] Analogously, a question of interest is reasoning in a time space, i.e. how far players look ahead. While a rational player should infinitely ahead (in an infinitely repeated game or a game with stochastic ending), Orland and Roos (2013) find that agents look ahead not more than four periods in an individual-decision making problem of price-setting. Woodford (2018) analyzes the effects of monetary policy with agents that only look ahead a finite number of periods and a value function learned from past experience to assess their situation after that finite number of periods.

and the left-hand neighbor. Subjects' payoff is such that they are most happy if they have at least one neighbor of the same type. In the first treatment, subjects move sequentially and without cost, i.e. one subject starts deciding upon her position in the circle, then the next one follows etc. Even after the circle is complete, a subject can move between two other subjects. In the second treatment, subjects are required to move simultaneously so that they make decisions about moving at the same time. The results of both treatments are identical to Schelling's prediction, as they both end up in the segregated equilibrium. In a later paper, Benito et al. (2015) extend Schelling's model by incorporating moving costs. Both the theoretical and the experimental results show that there is not full segregation in equilibrium. They also note that the degree to which players act strategically increases considerably with the moving costs.

4.5.3 Mixed Equilibrium Games in Lab and Field

Crawford and Iriberri (2007) explain the experimental results of the "hide and seek"-game by Rubinstein and Tversky (1993; "RT") and Rubinstein et al. (1996; "RTH"). Their experiment works as follows: Subjects were divided into hiders and seekers. The hider puts a prize in one of four boxes arranged in a row. Those boxes are marked by A, B, A, A. The seeker's task is to find the prize, only being allowed to open one box. This is similar to a matching pennies game, since one player wins by matching the opponent's choice while the opponent wins by mismatching. The A–B labeling is important to account for a "naturally occurring landscape" as in a hide-and-seek game but represents a tractable setting. It is argued that "B" is salient in the spirit of Thomas Schelling (1960). It is also argued that the two "end A" locations are salient in the sense of Christenfeld (1995). Rubinstein and Tversky (1993) argue that due to these two saliences render the remaining location, "central A", "the least salient place." Rubinstein et al. (1996) verify this choice as the modal behavior for both hiders and seekers, and was even more prevalent for seekers than hiders. Crawford and Iriberri (2007) propose a level-k explanation for this choice behavior. Level 0 is constrained to salient choices; level-1 hiders then best responds to that by using the non-salient location, "central A". While level-1 seekers match the anticipated behavior of level-0 hiders, level-2 seekers choose "central A" with probability 1 etc. Thus, the equal mixing in Nash-equilibrium is not observed. These games have not been played over long time horizon.

Field Experiments

Chiappori et al. (2002) note that evidence on players mixing strategies is based almost exclusively on laboratory experiments and express several criticisms on that methodology: first, laboratory experiments are "simplistic, artificial setting[s]"; second, stakes in the form of payoffs are low in laboratory experiments; third, subjects have other preferences than maximizing their payoffs such as "attempting to avoid looking foolish". It has also been argued that laboratory experiments contain too few rounds for subjects to gain significant experience (Palacios-Huerta, 2003). These criticisms have been responded by new directions.

Thus, papers like Walker and Wooders (2001), Chiappori et al. (2002), Palacios-Huerta (2003), Apesteguia and Palacios-Huerta (2010) propose the use of sports data that represent "generalized" matching penny games, i.e. where the Nash equilibrium is not necessarily (0.5, 0.5). Walker and Wooders analyze data on serves to start rallies in Grand Slam tennis tournaments, Chiappori et al. (2002) use the full sample of penalty kicks in the French and Italian elite soccer leagues over a period of three years, and Palacios-Huerta analyzes data on major professional soccer games in Spain, Italy, England and other countries. All three studies find the data to be consistent with Nash equilibrium. Apesteguia and Palacios-Huerta (2010) use data on penalty kicks in professional soccer tournaments as a natural experiment to test the hypothesis that agents make use of the first-mover advantage, since the team starting with penalty kicks is determined by a coin toss of the referee. The authors return a positive answer, showing that subjects are aware of the advantage of going first and rationally respond to it by systematically choosing to kick first when given the chance to choose the order. In contrast, in multi-period laboratory experiments with a mixed-strategy equilibrium, reinforcement models explain behavior over time better than the Nash equilibrium (see e.g. Erev and Roth, 1998).

Multi-Dimensional Iterative Reasoning: Colonel Blotto Game

This mixed equilibrium game by Arad and Rubinstein (2012b) is the most complicated situation we present here. Two players have to divide a given number of points (troops) into a specified number, e.g. 6, of battlefields. The person who has filled a field with more points than the opponent wins a fixed prize; a payoff of zero is awarded in cases of ties. The experiment is either run in the laboratory with students or conducted online with many participants as a tournament version of the Colonel Blotto Game. Three different dimensions of decision-making are involved in the multi-dimensional iterative reasoning extracted from the actual decisions: 1. In how many fields to concentrate, i.e. to invest most of the troops; 2. Whether to fill "disregarded" fields with one or two troops. 3. Into which fields to assign a large number troops (first, middle or last location). Separately, in each dimension, L0 chooses the intuitive category (e.g. an equal amount of troops); L1 chooses a category that properly responds to L0 within the dimension, etc. Despite the 250 million possible strategies, nine strategies were chosen by about 30% of the subjects. Another astonishing result is that the winners of both subject pools chose a distribution of (2, 31, 31, 31, 23, 2) across the six fields. The experimental results thus clearly document that patterns occur.

5 ELICITATION METHODS

In the previous section we presented a range of structural changes to the original Beauty Contest game to explain resulting behavioral heterogeneity and discrepancy with the theoretical benchmarks of mostly homogeneous solutions. Here we present other methods to create variation in behavior, which are not related to structural changes. This can be, for example, through inducing time pressure, i.e. giving subject limited time, instead of unlimited decision time.

Another important aspect in understanding the general patterns of behavior is to know the reasoning processes behind a choice. When classifying outcomes into certain categories, the same outcome can be created through various channels of cognitive procedures or emotional reactions. Consider two examples of choice: The first one is an equal split in an ultimatum game. Is the person acting according to some fairness principles or does he act strategically as he anticipates that the opponent expects a fair outcome? Secondly, how can one distinguish whether a subject plays a random choice, seemingly unrelated to the rules of the game, and a deliberate choice? For example, a subject in the $2/3$ average Beauty Contest game chooses 22, classified as a level 2 choice. However, her written explanation states "this is my age". By coincidence, this gives a high chance of winning when playing with undergraduate students. In the following, we present some elicitation methods which help reveal underlying thought processes and other ideas to induce behavioral changes.

5.1 STRATEGY (VECTOR) METHOD

One of the earliest and most used methods to systematically reveal reasoning processes behind choices in experimental economics has been the so-called strategy method. Selten (1967) introduces it for experiments with a two-step procedure. In the first step, subjects acquire knowledge about the game by playing it in the standard or direct response form reacting to actions of others. In a second step, subjects have to construct strategies with actions for each possible situation (information sets) they could possibly reach. The experimenter then calculates the payoffs according to a tournament structure. Each player's strategy is matched with all other strategies, e.g. in a two-person Cournot game there are thus $n(n-1)$ interactions for each player who is not matched with herself. Then the average payoff for all interactions are calculated.

There are only few experimental papers that utilize this two-step method, as originally introduced, e.g., Selten et al. (1997), and Keser (1996) in oligopoly games with cost asymmetries and demand inertia. Selten et al. (1997) find that subjects do not maximize profits based on expectation formation or backward induction in a 100-period Cournot game. Instead, in the initial periods of a programmed strategy, a common pattern among subjects is a formulation of a goal, an ideal point, which is a production outcome pair based on some fairness considerations. The ideal point can be interpreted as a reference point. Over time, the program describes a "measure for measure", that is, a move by the opponent towards one's own ideal point is responded by one's own move towards the ideal point, while a move away from the ideal point is followed by also moving away.

Mitzkewitz and Nagel (1993; MN) implement the strategy method without the direct response mode, for an experiment on ultimatum games with incomplete formation of the pie on the side of the responders. They pondered that these games are simple enough to understand without the first step. We describe this game in more

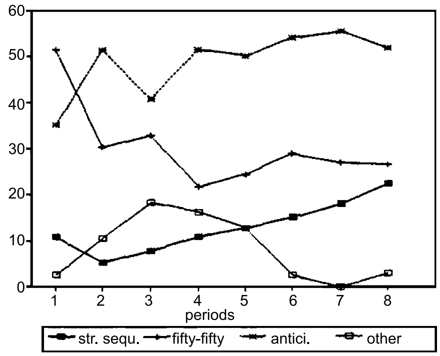

FIGURE 15

Types of (strategic) offer behavior in ultimatum offer games; y-axis denotes relative frequency of each type. str. sequ. = sequential equilibrium strategy of accepting all or all positive offers; fifty-fifty: proposal to split pie 50–50 and near 50–50; antici = anticipation strategy based on forming beliefs about an anticipated level of acceptance of the opponent. Source: Mitzkewitz and Nagel (1993).

detail as the rules allow for constructions of several reference points of the level-k model. The pie size is determined by rolling of a six-sided die. Proposers know the amount and specify an offer in offer games (demands, in demand games) for each possible amount. Responders respond to all possible offers (demands), not knowing the specific pie. Choices are from [0, 0.5, 1..., pie size]. With this method (see e.g. survey by Brandts and Charness, 2011), the researchers obtain information even at nodes typically not reached, for example low offers.

MN offer a three-step descriptive theory, resembling level-k. Let us exemplify this for the offer game. In a first step, a naive proposer, ignoring the information structure, offers half of the pie, assuming that the responder will accept half of the pie. A more sophisticated player assumes as a reference point half of the expected pie (3.5/2) in the offer games, and thus ponders to best respond with a rounded offer of 1.5 or 2 (second step). In a third step, offers are rounded downwards, when the offer is greater than half of the actual pie. This is one of the so-called anticipation strategies.

MN classify behavior according to equal split, anticipation strategies and sequential equilibrium strategies, allowing also for small deviations from these strategies. Equal split is the modal choice observed in the first period but this decays over time, as proposers gain a better understanding of the informational advantage for themselves. This decrease is not observed in complete information games, implying that the informational structure plays an additional role to determine fairness norms. Thus, as in level-k, beliefs do not need to be consistent, but instead learning mechanisms might converge to equilibrium strategies. Fig. 15 reflects the proportion of different reasoning patterns over time in an offer game.

Fischbacher et al. (2001) use the strategy method to understand conditional giving in public good games (see also Arifovic and Duffy, 2018). This also allows to understand the (slow) decrease of average giving in those games over time.

Stahl and Wilson (1994, 1995) construct several 3×3 normal-form games with dominance-solvable solutions. Each subject has to state an action and belief distribution of other players' actions for all games, without any feedback between the games. Given this strategy vector method, they classify subjects according to a level-k model which includes that a level-k player gives a best response to a distribution of lower levels (0 to $k - 1$), using also maximum likelihood estimations.

Risk Elicitation Method

Holt and Laury (2002) also use a kind of strategy vector method, a so-called list elicitation, to measure risk attitudes of subjects. In an individual decision-making task, subjects are confronted with a list of two lottery alternatives in each row (or one sure payoff increasing from row to row compared to a fixed lottery; see Heinemann et al., 2009). The rows are ordered in such a way that subjects should reveal their risk attitudes by choosing lotteries from one column for several rows and then switch once to the column of the other lotteries. At the switching point, the subject reveals the equivalence between the two lotteries (the certainty equivalence, in case of sure payoffs). This method produces a heterogeneity measure of risk attitudes. Typically there are few risk-loving subjects, while most show risk-averse attitudes. Since the beginning of this century, this kind of risk elicitation has been linked to behavior in games to analyze the relationship between strategic uncertainty and risk attitudes of a single subject (see, e.g., the literature cited in Heinemann et al., 2009).

5.2 RESPONSE TIME (RT)

Measuring response time is probably one of the least inflicting methods to understand features of the reasoning procedures, since subjects cannot notice that the analyst collects these data. It has been widely used in psychology to understand the difference between deliberation in thought processes and random, spontaneous, or intuitive choices. Using computers to solicit behavior either in the computer laboratory or with online platforms like Qualtrics®, the researcher can obtain reaction times for each choice accurately up to 50 milliseconds. Rubinstein (2007) analyzes reaction time (RT) in several experiments, among others the Beauty Contest games, ultimatum games, etc. For the Beauty Contest game data he finds that the RT of those who choose 22 (4%) was 157 seconds while the RT among the 8 who choose 20, 21, 23, 24, 25 was only 80 seconds. He concludes that "[t]his must mean that there is little in common between the choice of 22 and the rest of the category which Nagel called 'Step 2'." However, a choice of 20 can be made faster, as the subject, e.g., might jump from 50 to 30 to 20 without any precise mental calculations while a choice of 22.22 requires higher calculation time. Yet, both 22.22 and 20 might involve level-k thinking, one with precise and the other one with rough calculations. These discrepancies of data interpretations call for a collaboration between researchers, using different methods to arrive at a more conclusive understanding of actual behavior. For a survey on experiments with RT see Brañas-Garza et al. (2017).

5.3 MOUSELAB, EYE-TRACKING

An increasing literature uncovers the reasoning processes through various (new) choice elicitation techniques. Costa-Gomes et al. (2001) use mouselab tracking techniques by hiding the information of payoffs in normal form games behind boxes on the computer screens which subjects need to click to receive the information. The authors specify possible look-up-ex-ante rules. For instance, for level 1 reasoning the player only needs to look up her own choices, while level 2 reasoning requires lookups of payoffs from the other player. Roos and Luhan (2013) study whether subjects use all available information in a more complicated setup, where information is only revealed if subjects "buy" it. They find that 75% of the expectations formed in their setup did not use all information available. Brocas et al. (2014) find evidence for level-k thinking in betting games when using this technique. The eye-tracking results of Müller and Schwieren (2011) suggest that subjects in Beauty Contest experiments actually think (or look) more steps ahead than their chosen numbers reveal. Chen and Krajbich (2017) implement eye-trackers while subjects play two-person Beauty Contest games with tournament payoffs. They are in particular interested when subjects realize that 0 is an optimal strategy. The measure of pupil dilation and eye movements are clearly distinguishable between those who switch to the winning strategy (epiphany learners) as compared with subjects who do not learn the winning strategy.

5.4 CONTINUOUS TIME DECISION-MAKING

Whether a choice is determined randomly or with deliberation can be elicited by letting the subjects continuously choose within a given time interval. Agranov et al. (2015) introduce this method in a basic Beauty Contest game and pay subjects according to their choices in a particular time period, chosen at the end, unknown to the subjects ex ante. They find a clear distinction between those who do not understand the rules of the game, identified by random walk behavior, and those who discover either level 1 or level 2 choices after some learning has occurred.

5.5 WRITTEN COMMENTS, CHATS, AND OTHER COMMUNICATION TOOLS, TEAM DECISIONS

Nagel (1993), Bosch et al. (2002), and many others nowadays ask subjects to explain their choices in a written form during or after the experiments. These comments reveal especially random choices or very clear-cut reasoning procedures like choosing $50 \cdot p^k$ in Beauty Contest games. Bosch et al. classify over one thousand comments sent to the experimenters in collaboration with three different newspapers. One interesting finding is that those who comment on the equilibrium (80% of those equilibrium commentators) typically choose numbers between 0 and 10, thus remaining far from the winning choice in these newspaper experiments around 14 to 17.

Burchardi and Penczynski (2014) allow written communication within two-person teams, playing the basic Beauty Contest game and analyze the reasoning procedure related to level k. Sbriglia (2008) provides subjects with comments written by the winners of previous rounds, giving advice on how to choose in Beauty Contest games. This increases the speed of convergence to equilibrium.

The most commonly used communication tool is through the direct application of the so-called cheap talk where one or all subjects send messages related to the action space, e.g. I will choose “action” A (see e.g. Charness, 2000 for stag-hunt games and PD games). In the former, one-sided messages achieve coordination but in PD games messages are not truthful and thus behavior does typically not achieve efficiency.

5.6 ASKING FOR GUESSES

Haruvy (2002) proposes direct elicitation of beliefs to support level-k behavior, additionally to analyzing the actions of the same players. He discovered that explicit beliefs exhibit level-k patterns. Arjona et al. (2017) ask one set of subjects for their choices and another set about their guesses about opponents' distribution of choices in Arad and Rubinstein (2012a) modified two-person 11–20 games, in which a bonus payment is given to the player whose choice is by $R = 3$ steps (AR use $R = 1$) lower than his opponent (e.g. one chooses 20 and the other chooses 17 and therefore receives a bonus addition to the payment of 17). They distinguish more clearly between level-k choices and noisy choices than when $R = 1$ as in AR. About 74% of the guesses and the actual choices can be classified as level 0 (choice 20), level 1 (choice 17), level 2 (choice 14), or level 3 (choice 11).

5.7 PERSONAL CHARACTERISTICS AND TRAINING

A question of interest is to what extent intrinsic characteristics play a role. The subject pool can be quite heterogeneous and thus might create heterogeneous experimental outcomes. Subjects can be inexperienced or experts, students or professionals. Some studies concentrate on children as subjects. Grueneisen et al. (2015) test whether six-year olds can form higher-order beliefs in strategic interaction with peers and return a positive answer, in contrast to previous literature.

Clearly, decision-making very much depends on the context of the environment but also on personal characteristics of a subject. These can be objective measures such as age, gender, training with math, but also cognitive abilities which need to be elicited through various (psychological) tests, such as IQ tests. These characteristics might also influence the reaction towards others induced by expectations about other agents' characteristics.

The careful study on gender has entered experimental economics in the last 15 years (see a survey by Niederle, 2016). Whether there are gender differences related to level-k in BC games depends very much on the details of the experimental design. Cubel and Sanchez-Pages (2017) find that females display less depth of reasoning than males without financial incentives, but no differences prevail in payment treatments. Yet, making mixed gender composition salient induces lower choices for females as compared to males. The reason is that in general women believe that they are better in reasoning in such types of games than men, and correspondingly males believes this as well, which leads them to use less depth of reasoning. The authors relate their finding to stereotype threat (Steele, 1997), which states that members of negatively stereotyped groups indeed choose worse in fear of confirming the stereotype.

Training in certain cognitive realms like chess players (Bühren and Frank, 2012) shows no difference in behavior in BC-experiments, while economists obviously play closer to equilibrium (Bosch-Domènech et al., 2002). Dickinson and McElroy (2010) show that sleep deprivation when playing the BC results in diverting even further from equilibrium play. The authors call up the subjects at several times of the day, also at night. Gill and Prowse (2016) find that subjects with a higher cognitive ability, measured by the Raven test, choose numbers closer to the equilibrium in Beauty Contest games over time, but not in the initial periods. They also investigate the connections between level-k behavior and non-cognitive abilities. Furthermore, cognitive ability improves reasoning in the BC games (Burnham et al., 2009; Brañas-Garza et al., 2012). Additionally, cognitively more able individuals who believe that others are more clever than themselves revise their choices in BC games downwards. This is not the case for those who have lower scores in cognitive ability (Fehr and Huck, 2016).

Bosch-Rosa et al. (2016, see also above) devised the one person BC game, together with other cognitive measures (Raven test) and the results in the original Beauty Contest game to construct an index to discriminate between "high sophistication" and "low sophistication" subjects. High types among themselves (without knowing this grouping) create no bubbles in subsequent asset markets, unlike when only low types interact with each other (see also Arifovic and Duffy, 2018). Levine et al. (2017) also combine several IQ tests, BC-games, and "bubble" experiments, randomly distributing the subjects to different groups. They show that subjects with high strategic IQ (measured by Beauty Contest games) and cognitive intelligence (measured by cognitive reflection test (CRT) and other tests) result in higher earnings in subsequent asset markets as compared to players of the type "low" within these measures.

5.8 NEUROECONOMICS

Since the early 2000s, biological data were gathered as a new source of explaining heterogeneity in behavior through collaborations with neuroscientists (see survey by Camerer et al., 2016). This leads to new design challenges when implementing technical instruments like fMRI (frequency magnetic resonance imaging) scanners measuring brain activity (via blood oxygen levels, known as BOLD signals, which are considered a proxy for cellular activity) and eye-tracking (following eye movements and pupil dilation etc.).

In experimental economics typically "between"-subject comparison is implemented, i.e., a design where one group of subjects participates only in one particular setup and their behavior is compared with the one of subjects in another setup. The economic researcher is typically interested in seeing the isolated consequences due to a single exogenous change in the design. The situations, interactions or decisions are mostly repeated, to understand the difference in perception of new situations and learning over time.

In contrast to economics, in neuroscience, multiple observations from various parameter changes or setups (e.g. playing against computer or playing against humans) are required from the same subject. The reason for these differences in design is that brain activity produces high variance also within a single subject and not only between subjects, and imaging requires contrasting brain activities through different tasks for the statistical analysis. Furthermore, the usage of individualized technological instruments like eye-trackers, skin conductances typically cannot be used for more than a few, or is restricted to even only to one subject, especially when using fMRI. There have been many more brain image studies about individual decision-making (e.g. risk attitudes, intertemporal choices, human well-being etc.) than interactive games (social preferences and strategic behavior we discussed in Sections 2 and 4).

In this section, we describe two studies, which on the one hand provide a biological foundation of level-k reasoning, and on the other hand show differences in brain activity when dividing subjects into high and low earners according to their realized payoffs. Coricelli and Nagel (2009) study behavior in Beauty Contest games using fMRI brain imaging. All subjects are invited one by one to participate in various games with different b-parameters, playing every other period either against nine computer programs, who just choose randomly between 0 and 100, or against nine human subjects who are recorded the same way. There is no feedback between rounds. They also solve calculation tasks related to the level-k calculations as $50 \cdot (2/3)$ or $50 \cdot (2/3)^2$. This is all known to the participants. All subjects behave as level 1 when playing against the randomized computer programs. However, with human opponents most choices are still related to level-1 reasoning, while about 1/3 of the subjects reveal level-2 or higher levels. Brain activity of these two different reasoning types are well distinguishable with higher activity in the medial prefrontal cortex (mPFC) for level-2 players, a theory of mind area (i.e. thinking what others are thinking) and rostral anterior cingulate cortex (rACC) for level-1 players which is related to self referential thinking (see Fig. 16). Yet, there are also brain regions which produce similar activities for high and low reasoning types, as in the superior temporal sulcus (STS), posterior cingulate cortex, and bilateral temporo-parietal junction (TPJ), which are areas responsible for general planning. When playing against the computer, no differential activity in these areas was noted, but only in the calculation area as angular gyrus. This is similar as in the simple calculation tasks we added at the end within the scanner.

Concluding this study, even though the level-k model is specified as a mental model of calculation $50 \cdot (2/3)^k$, the interpretation of putting oneself into the shoe of others can be now supported with brain activity. Furthermore, newspaper participants, who write comments like “I think that the average will be 50 therefore I will choose 33, but others will think alike and therefore I choose 25”, reveal a reasoning process beyond mental calculations with theory of mind ideas involved, which is visualized in the mPFC.

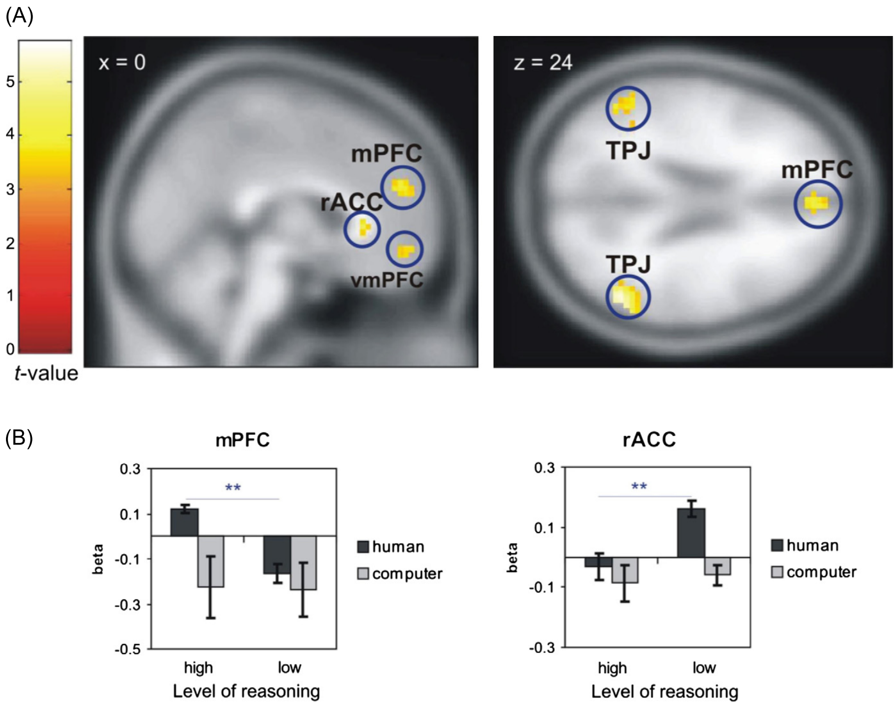

FIGURE 16

Coricelli and Nagel (2009): Patterns of behavior and brain activity for low and high levels of reasoning. The bar charts show the differences of BOLD signals for low vs high reasoning types playing either against humans or computer programs.

The second study is by Smith et al. (2014), who implement a bubble experiment as in Smith et al. (1988, also mentioned in Section 2) with a risky asset and risk-free cash, interacting for 50 periods. Two or three subjects are placed in an fMRI machine and 9–21 subjects as usual in front of the computer. The authors classify players into low, medium and high earners showing differences in trading behavior and brain activity between the three different earning types. Fig. 17 exemplifies heterogeneity in terms of trading behavior and brain activity: average right anterior insula activity of low earners oscillates around 0, while high earners on average show an activity peak that coincides with the beginning of selling units 5–10 periods before the peak of the bubble. The authors interpret this result with high earners producing some physical discomfort and therefore exit the asset market before the crash occurs.

6 DISCUSSION

In the 1950s, the Air force quickly acknowledged the heterogeneity of its pilots and demanded that cockpits should fit pilots falling within the 5 to 95 percent range on

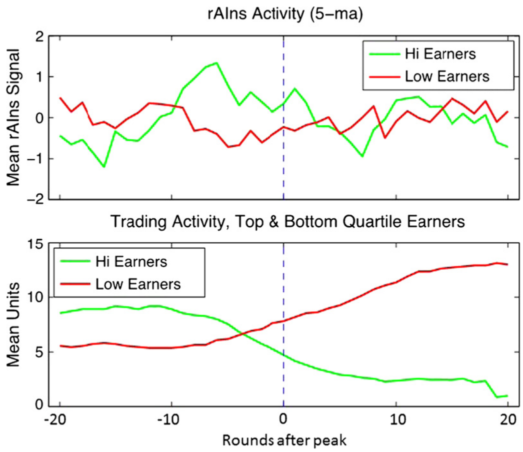

FIGURE 17

Smith et al. (2014): Average right anterior insula activity in high earners and low earners shows that low earner activity fluctuates around 0, whereas high earner activity shows a peak that coincides with the beginning of the sell-off of units shown in (A) (5–10 periods before the price peak).

each dimension (Rose, 2016, p. 9). Adjustable seats were introduced, being an innovation at the time. How to fit economic theory to heterogeneity is less obvious, but we hope to have shown that experimental economics contributes to the understanding of subjects' behavior in the relevant 5 to 95 percent ranges.

To understand metaphorically the rich amount of regularity we have explored as a contrast to the mostly homogeneous rationality benchmark, we introduce the tuning fork (see Nagel et al., 2016, p. 79; Nagel, 2018). It has most of the vibrational energy at the fundamental frequency. In contrast humans with all their shortcomings and rich set of emotions would then represent the instruments, each with different overtones, which however, also have their regularities, as expressed with the level-k model. This helps establish a more positive view of human interaction, as in an orchestra with its diversity, and the tuning fork as an important coordination device. But no instrument would like to be only a tuning fork, pure and poor.

In this chapter, the models, representing economic situations, draw on a universal basic structure, the Beauty Contest game, as a canonical structure to create several archetypal games in (experimental) economics that at first seem disconnected. Identifying a particular game or function within a general structure can give the analyst a hunch about how individual subjects in the laboratory would empirically behave. An encompassing function containing a best response to others' behavior and further exogenous parameters seems a natural candidate: a reasonable or thoughtful subject forms possibly sophisticated beliefs in her mind about the play of a single opponent, or group of opponents and other external features. In such a mind – an aggregation of choices might be represented by an average or reference point. A second step is thereafter to best reply towards to this belief and so on, to maybe reach equilibrium.

Herbert Simon (1996, p. 34ff) has pondered about such a Cournot (1838) equilibrium tatonnement process "each firm, with limited cleverness, formed an expectation of its competitor's reaction to its actions, but that each 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). Von Neumann started from the other end, the equilibrium, asking "how is the absolutely selfish 'homo economicus' going to act under given external circumstances?" The key answer, more parsimonious than maybe first thought, has been given by the Keynesian Beauty Contest metaphor with several degrees of reasoning. While the game had been widely used in the macroeconomic literature (see, e.g., discussion by Nagel et al., 2017), the reasoning process has been ignored for almost 60 years. With our chapter we hope to contribute to a behavioral (micro-)foundation of macroeconomics, for which Keynes has made some initial suggestions.

The level-k model, an empirically validated model in the spirit of Keynes, has been a successful descriptive (boundedly) rational model to account for these kinds of observations and questions, proposing limited reasoning. The so-called black box has been opened but restrained at the same time, bridging the gap between fully rational reasoning and beliefs, and random behavior. Yet, in describing heterogeneity of subjects' all types of reasoning can be found seated side by side. Knowing the presence of different types might identify the mistakes of (aggregate) behavior or failures of coordination to provide tools for improving either the institutional environment or the play of a single agent.

While we consider Beauty Contest games only with linear aggregation rules, we hope this can be seen as a starting point and potential future research efforts can be directed towards how the definition of Beauty Contest game can be extended, restricted or refined. Still, one will also need to understand how interactive situations not in this class can be identified and potentially generalized or systematized.

In a similar spirit, level-k has at least two limitations: first, it represents a (descriptive) model being silent about how the distribution of types emerges. Yet, if the researcher knows the subject pool, the result can be more easily predicted, as there is an astonishing robustness of results depending on the population. Second, subjects do not always have enough information or knowledge about the best response function to engage in reasoning about other players' reasoning depth. In those cases, focal or reference points, random behavior, moral laws, heuristics or intuition are guiding principles without the need of higher-order reasoning. Notably, reference points as the key elements of the level-k model are in need of empirical validation. Here a link to the rich studies and heuristics from individual decision making might become of help.

Over time adaptive processes with more or less information feedbacks provide additional patterns, which are covert in more detail in the chapter by Arifovic and Duffy. The authors discuss behavior, for example, in NK-models that fall into our classification of BC-games but are more complicated. Finally, the knowledge of the environment is often either not known or too complicated to allow reasoning about other subjects' reasoning.

There is yet, another criticism about the content presented in this chapter and maybe about the current state of some parts of experimental economics. Returning to our initial example of the air force, the pilots owed the invention of the adjustable seat to a visionary scientist who detected the extent of heterogeneity. This was possible because he measured painstakingly many dimensions of pilots' bodies. Here, we presented different situations with focus on single variable decisions. As a contrast, Reinhard Selten asked for experimental studies of complicated situations with many different decision variable and actors. He stated many examples from the very beginning of his career in economics in the 1950s such as oligopolies (see Abbink and Brandts, 2010; Nagel et al., 2016). Another example of a more complicated structure is also Arad and Rubinstein (2012b) which we presented above. Such studies will reveal further interesting heuristics (e.g. how a single subject copes with many decision variables; see Selten et al., 2012), analogously to the adjustable seat situation. We hope that further inquiries in this vein result in fruitful collaborations among experimental economists, akin to computer scientists who provide new tools for exploring and understanding heterogeneity in big data science.

REFERENCES

Abbink, K., Brandts, J., 2010. Drei Oligopolexperimente. In: The Selten School of Behavioral Economics. Springer, Berlin, Heidelberg, pp. 53–72.

Adam, K., 2007. Experimental evidence on the persistence of output and inflation. The Economic Journal 117 (520), 603–636.

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.

Agranov, Marina, Potamites, Elizabeth, Schotter, Andrew, Tergiman, Chloe, 2012. Beliefs and endogenous cognitive levels: an experimental study. Games and Economic Behavior 75 (2), 449–463.

Alaoui, L., Penta, A., 2016a. Endogenous depth of reasoning. The Review of Economic Studies 83 (4), 1297–1333.

Alaoui, L., Penta, A., 2016b. Cost–benefit analysis in reasoning. Mimeo.

Alaoui, L., Penta, A., 2017. Reasoning About Others' Reasoning. Barcelona GSE Working Paper 1003.

Amir, R., 2005. Supermodularity and complementarity in economics: an elementary survey. Southern Economic Journal 71 (3), 636–660.

Angeletos, G.-M., Hellwig, C., Pavan, A., 2007. Dynamic global games of regime change: learning, multiplicity, and the timing of attacks. Econometrica 75, 711–756.

Angeletos, G.-M., La'O, J., 2010. Noisy business cycles. In: NBER Macroeconomics Annual 2009, vol. 24. University of Chicago Press, pp. 319–378.

Angeletos, G.-M., Lian, C., 2016. Incomplete information in macroeconomics. In: Handbook of Macroeconomics, vol. 2.

Angeletos, G.-M., Werning, I., 2006. Crises and prices: information aggregation, multiplicity, and volatility. The American Economic Review 96 (5), 1720–1736.

Apesteguia, J., Palacios-Huerta, I., 2010. Psychological pressure in competitive environments: evidence from a randomized natural experiment. The American Economic Review 100 (5), 2548–2564.

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

Arad, Ayala, Rubinstein, Ariel, 2012b. Multi-dimensional iterative reasoning in action: the case of the Colonel Blotto game. Journal of Economic Behavior & Organization 84 (2), 571–585.

Arifovic, J., Duffy, J., 2018. Heterogeneous agent modelling: experimental evidence. In: Handbook of Computational Economics, vol. 4. Elsevier, pp. 491–540 (this Handbook).

Arifovic, J., Hommes, C., Kopanyi-Peuker, A., Salle, I., 2018. Are sunspots effective in a big crowd? Evidence from a large-scale bank run experiment. Mimeo.

Arifovic, J., Petersen, L., 2017. Stabilizing expectations at the zero lower bound: experimental evidence. Journal of Economic Dynamics and Control 82, 21–43.

Arjona, D.R., Fragiadakis, D.E., Kovaliukaite, A., 2017. The predictive success of theories of strategic thinking: a non-parametric evaluation.

Assenza, T., Heemeijer, P., Hommes, C., Massaro, D., 2014. Managing Self-Organization of Expectations Through Monetary Policy: A Macro Experiment. CeN-DEF Working Paper 14-07.

Aumann, R.J., 1974. Subjectivity and correlation in randomized strategies. Journal of Mathematical Economics 1, 67–96.

Aumann, R.J., 1987. Correlated equilibrium as an expression of Bayesian rationality. Econometrica 55, 1–18.

Avery, C., Kagel, J.H., 1997. Second-price auctions with asymmetric payoffs: an experimental investigation. Journal of Economics & Management Strategy 6 (3), 573–603.

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

Banerjee, A.V., Duflo, E., 2009. The experimental approach to development economics. Annual Review of Economics 1 (1), 151–178.

Banerjee, A.V., Duflo, E. (Eds.), 2018. Handbook of Economic Field Experiments, vol. 2, pp. 3–654.

Banerjee, A., Weibull, J.W., Binmore, K., 1996. Evolution and rationality: some recent game-theoretic results. In: Economics in a Changing World. Palgrave Macmillan, London, pp. 90–117.

Bao, T., Hennequin, M., Hommes, C., Massaro, D., 2016. Coordination on Bubbles in Large-Group Asset Pricing Experiments. Working paper. University of Amsterdam.

Bao, T., Hommes, C., 2017. When Speculators Meet Constructors: Positive and Negative Feedback in Experimental Housing Markets. CeNDEF Working Paper. University of Amsterdam.

Bao, T., Hommes, C., Sonnemans, J., Tuinstra, J., 2012. Individual expectations, limited rationality and aggregate outcomes. Journal of Economic Dynamics and Control 36, 1101–1120.

Baroni, M., Maguire, S., Drabkin, W., 1983. The Concept of Musical Grammar. Music Analysis 2, 175–208.

Bayona, A., Brandts, J., Vives, X., 2016. Supply Function Competition, Private Information, and Market Power: A Laboratory Study. CEPR Discussion Paper 11378. Center for Economic and Policy Research (CEPR), London.

Beja, A., Goldman, M.B., 2015. On the dynamic behavior of prices in disequilibrium. The Journal of Finance 35, 235–248.

Benhabib, J., Bühren, C., Duffy, J., Nagel, R., 2018a. Strategic nudging. Mimeo.

Benhabib, J., Duffy, J., Nagel, R., 2018b. De-framing the rules to (de)-anchor beliefs in beauty contest games. Mimeo.

Benhabib, J., Wang, P., Wen, Y., 2015. Sentiments and aggregate demand fluctuations. Econometrica 83, 549–585.

Benito, J., Branas, P., Hernandez, P., Sanchis, J., 2011. Sequential vs. simultaneous Schelling models. The Journal of Conflict Resolution 55 (1), 33–59.

Benito, J., Branas, P., Hernandez, P., Sanchis, J., 2015. Strategic behavior in Schelling dynamics: a new result and experimental evidence. Journal of Behavioral and Experimental Economics 57, 134–147.

Bergemann, D., Morris, S., 2013. Robust predictions in games with incomplete information. Econometrica 81 (4), 1251–1308.

Bolton, G., Ockenfels, A., 2000. ERC: a theory of equity, reciprocity, and competition. The American Economic Review 90, 166–193.

Bordalo, Pedro, Gennaioli, Nicola, Shleifer, Andrei, 2012. Salience theory of choice under risk. The Quarterly Journal of Economics 127 (3), 1243–1285.

Bosch-Domènech, A., Montalvo, J.G., Nagel, R., Satorra, A., 2002. One, two, (three), infinity, ...: newspaper and lab beauty-contest experiments. The 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., 2016. Cognitive bubbles. Experimental Economics. Forthcoming.

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

Brañas-Garza, P., Meloso, D., Miller, L., 2017. Strategic risk and response time across games. International Journal of Game Theory 46 (2), 511–523.

Branch, W., McGough, B., 2009. A new Keynesian model with heterogeneous expectations. Journal of Economic Dynamics & Control 33, 1036–1051.

Brandenburger, A., Danieli, A., Friedenberg, A., 2017. How many orders do player's reason? An observational challenge and solution. Mimeo.

Brandts, J., Charness, 2011. The strategy versus the direct-response method: a first survey of experimental comparisons. Experimental Economics 14, 375–398.

Brandts, J., Potters, J., 2017. Experimental industrial organization. In: Handbook of Game Theory and Industrial Organization. Edward Elgar Publishing. Forthcoming.

Breaban, A., Noussair, C.N., 2015. Trader characteristics and fundamental value trajectories in an asset market experiment. Journal of Behavioral and Experimental Finance 8, 1–17.

Breitmoser, Y., 2012. Strategic reasoning in p-beauty contests. Games and Economic Behavior 75, 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.

Brock, W.A., Hommes, C., 1998. Heterogeneous beliefs and routes to chaos in a simple asset pricing model. Journal of Economic Dynamics & Control 22 (8–9), 1235–1274.

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 & Organization 72 (1), 171–175.

Bryant, J., 1983. A simple rational expectations Keynes-type model.. The Quarterly Journal of Economics 98, 525–528.

Cabrales, A., Nagel, R., Armenter, R., 2007. Equilibrium selection through incomplete information in coordination games: an experimental study. Experimental Economics 10, 221–234.

Camerer, C., 1995. Individual Decision Making. Princeton University Press, Princeton, NJ.

Camerer, C.F., 2003. Behavior Game Theory: Experiments in Strategic Interaction. Princeton University Press, Princeton, NJ.

Camerer, C., Cohen, J., Fehr, E., Glimcher, P., Laibson, D., 2016. Neuroeconomics. In: Handbook of Experimental Economics, vol. 2.

Camerer, C., Fehr, E., 2006. When does "economic man" dominate social behavior? Science 311, 47.

Camerer, C., Ho, T.H., 1999. Experience-weighted attraction learning in normal-form games. Econometrica 67, 827–874.

Camerer, C., Ho, T.H., Chong, J.-K., 2002. Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. Journal of Economic Theory 104, 137–188.

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

Camerer, Colin F., Loewenstein, George, Weber, Martin, 1989. The curse of knowledge in economic settings. Journal of Political Economy 97 (5), 1232–1254.

Campbell, J.Y., Lo, A.W.C., MacKinlay, A.C., 1997. The Econometrics of Financial Markets. Princeton University Press, Princeton, NJ.

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

Charness, G., 2000. Self-serving cheap talk: a test of Aumann's conjecture. Games and Economic Behavior 33 (2), 177–194.

Charness, G., Rabin, M., 2002. Understanding social preferences with simple tests. The Quarterly Journal of Economics 117, 817–869.

Chen, W.J., Krajbich, I., 2017. Computational modeling of epiphany learning. Proceedings of the National Academy of Sciences 114 (18), 4637–4642.

Chen, Y., Konstan, J., 2018. Online field experiments: a selective survey of methods. Mimeo.

Chiappori, P.A., Levitt, S., Groseclose, T., 2002. Testing mixed-strategy equilibria when players are heterogeneous: the case of penalty kicks in soccer. The American Economic Review 92 (4), 1138–1151.

Chompsy, Noam, 1957. Syntactic Structures. Mouton, The Hague.

Chong, J.K., Ho, T.H., Camerer, C., 2016. A generalized cognitive hierarchy model of games. Games and Economic Behavior 99, 257–274.

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

Christenfeld, N., 1995. Choices from identical options. Psychological Science 6 (1), 50–55.

Cooper, R., 1999. Coordination Games. Complementarities and Macroeconomics. Cambridge University Press, Cambridge, UK.

Cooper, D., Kagel, J., 2016. Other-regarding preferences: a selective survey of experimental results. In: Kagel, J., Roth, A. (Eds.), Handbook of Experimental Economics, vol. 2.

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.

Cornand, C., 2006. Speculative attacks and informational structure: an experimental study. Review of International Economics 14, 797–817.

Corsetti, G., Guimaraes, B., Roubini, N., 2006. International lending of last resort and moral hazard: a model of IMF's catalytic finance. Journal of Monetary Economics 53, 441–471.

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.

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

Cournot, A.A., 1838. Recherches sur les principes mathématiques de la théorie des richesses par Augustin Cournot. L. Hachette.

Crawford, V., 1995. Adaptive dynamics in coordination games. Econometrica 65, 103–143.

Crawford, V., 2013. Boundedly rational versus optimization-based models of strategic thinking and learning in games. Journal of Economic Literature 51, 512–527.

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., 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., Pauzner, A., 2009. Behaviorally optimal auction design: examples and observations. Journal of the European Economic Association 7 (2–3), 377–387.

Cubel, M., Sanchez-Pages, S., 2017. Gender differences and stereotypes in strategic reasoning. The Economic Journal 601, 728–756.

Daniels, G.S., 1952. The Average Man? No. TN-WCRD-53-7. Air Force Aerospace Medical Research Lab, Wright-Patterson AFB, OH.

Davis, D.D., Holt, C.A., 1993. Experimental Economics. Princeton University Press.

De Long, J.B., Shleifer, A., Summers, L., Waldmann, R., 1990. Noise trader risk in financial markets. Journal of Political Economics 98 (4), 703–738.

Dhami, S., 2016. Foundations of Behavioral Economic Analysis. Oxford University Press.

Diamond, P.A., 1982. Aggregate demand management in search equilibrium. Journal of Political Economy 90, 881–894.
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.
Dixit, A., 2006. Thomas Schelling's contributions to game theory. Scandinavian Journal of Economics 108, 213–229.
Duffy, J., 2016. Macroeconomics: a survey of laboratory research. In: Kagel, J.H., Roth, A.E. (Eds.), Handbook of Experimental Economics, vol. 2. Princeton University Press.
Duffy, J., Nagel, R., 1997. On the robustness of behaviour in experimental 'beauty contest' games. The Economic Journal 107 (445), 1684–1700.
Duffy, J., Ochs, J., 2012. Equilibrium selection in entry games: an experimental study. Games and Economic Behavior 76, 97–116.
Duflo, E., 2006. Field Experiments in Development Economics. Econometric Society Monographs, vol. 42, p. 322.
Dufwenberg, M., Kirchsteiger, G., 2004. A theory of sequential reciprocity. Games and Economic Behavior 47, 268–296.
Eppel, T., Matheson, D., Miyamoto, J., Wu, G., Eriksen, S., 1992. Old and new roles for expected and generalized utility theories. In: Edwards, Ward (Ed.), Utility Theories: Measurements and Applications. Kluwer Academic Press, Boston, pp. 271–294.
Erev, I., Haruvy, E., 2016. Learning and the economics of small decisions. In: Kagel, J.H., Roth, A.E. (Eds.), The Handbook of Experimental Economics, vol. 2. Princeton University Press, pp. 638–716.
Erev, I., Roth, A., 1998. Predicting how people play in games: reinforcement learning in experimental games with unique, mixed strategy equilibria. The American Economic Review 88, 848–881.
Evans, G.W., Honkapohja, S., 2001. Learning and Expectations in Macroeconomics. Princeton University Press.
Eyster, E., Rabin, M., 2005. Cursed equilibrium. Econometrica 73 (5), 1623–1672.
Falk, A., Fischbacher, U., 2006. A theory of reciprocity. Games and Economic Behavior 54, 293–314.
Farhi, E., Werning, I., 2017. Monetary Policy, Bounded Rationality, and Incomplete Markets. No. w23281. National Bureau of Economic Research.
Fehr, E., Gachter, S., 2000. Cooperation and punishment in public goods experiments. The American Economic Review 90, 980–994.
Fehr, D., Heinemann, F., Llorente-Saguer, A., 2017. The Power of Sunspots: An Experimental Analysis. Working paper.
Fehr, Dietmar, Huck, Steffen, 2016. Who knows it is a game? On strategic awareness and cognitive ability. Experimental Economics 19, 713–726.
Fehr, E., Schmidt, K., 1999. A theory of fairness, competition and cooperation. The Quarterly Journal of Economics 114, 817–868.
Fehr, E., Tyran, J-F., 2001. Does money illusion matter? The American Economic Review 91, 1239–1262.
Fehr, E., Tyran, J.-F., 2007. Money illusion and coordination failure. Games and Economic Behavior 58, 246–268.
Fehr, E., Tyran, J.-R., 2008. Limited rationality and strategic interaction: the impact of the strategic environment on nominal inertia. Econometrica 76, 353–394.
Fischbacher, U., Gächter, S., Fehr, E., 2001. Are people conditionally cooperative? Evidence from a public goods experiment. Economics Letters 71 (3), 397–404.
Fragiadiakis, D., Kovaliukaite, A., Arjona, D.R., 2017. Testing Cognitive Hierarchy Assumptions. Working paper.
Fréchette, Guillaume R., 2016. Experimental economics across subject populations. In: Kagel, John H., Roth, Alvin E. (Eds.), The Handbook of Experimental Economics, vol. 2. Princeton University Press, pp. 435–480.
Friedenberg, A., Kets, W., Kneeland, T., 2015. Bounded reasoning: rationality or cognition. Mimeo.
Fudenberg, D., Levine, D., 1998. The Theory of Learning in Games. MIT Press, Cambridge and London.
Gale, D., 1995. Dynamic coordination games. Economic Theory 5 (1), 1–18.

Gale, D., Shapley, L.S., 1962. College admissions and the stability of marriage. The American Mathematical Monthly 69, 9–14.
Galí, J., 2008. Monetary Policy and the Business Cycle. Princeton University Press, Princeton, NJ.
García-Schmidt, M., Woodford, M., 2015. Are Low Interest Rates Deflationary? A Paradox of Perfect-Foresight Analysis. No. w21614. National Bureau of Economic Research.
Garratt, R., Wooders, J., 2010. Efficiency in second-price auctions: a new look at old data. Review of Industrial Organization 37, 42–50.
Georganas, S.C., Healyb, P.J., Weber, R.A., 2015. On the persistence of strategic sophistication. Journal of Economic Theory 159 (Part A), 369–400.
Georganas, S., Nagel, R., 2011. Auctions with toeholds: an experimental study of company takeovers. International Journal of Industrial Organization 29 (1), 34–45.
Giamattei, Marcus, 2017. Cold Turkey vs. Gradualism: Evidence on Disinflation Strategies from a Laboratory Experiment. Passauer Diskussionspapiere. Volkswirtschaftliche Reihe, No. V-67-15.
Giamattei, M., Lambsdorff, J. Graf, 2015. Balancing the current account – experimental evidence on underconsumption. Experimental Economics 18 (4), 679–696.
Gigerenzer, G., Todd, P.M., ABC Research Group, 1999. Simple Heuristics That Make Us Smart. Oxford University Press.
Gill, David, Prowse, Victoria, 2016. Cognitive ability and learning to play equilibrium: a level-k analysis. Journal of Political Economy 126 (4), 1619–1676.
Gintis, H., 2003. Strong reciprocity and human sociality. Journal of Theoretical Biology 206, 169.
Gneezy, U., Niederle, M., Rustichini, A., 2003. Performance in competitive environments: gender differences. The Quarterly Journal of Economics 118 (3), 1049–1074.
Gode, D., Sunder, S., 1993. Allocative efficiency of markets with zero-intelligence traders: market as a partial substitute for individual rationality. Journal of Political Economy 101, 119–137.
Goeree, J.K., Holt, C.A., 2004. A model of noisy introspection. Games and Economic Behavior 46 (2), 365–382.
Goeree, J.K., Holt, C.A., Palfrey, T.R., 2002. Quantal response equilibrium and overbidding in private-value auctions. Journal of Economic Theory 104 (1), 247–272.
Goeree, J.K., Holt, C.A., Palfrey, T.R., 2003. Risk averse behavior in generalized matching pennies games. Games and Economic Behavior 45 (1), 97–113.
Goeree, J.K., Holt, C.A., Palfrey, T.R., 2016. Response Equilibrium: A Statistical Theory of Games. Princeton University Press, Princeton.
Goeree, J.K., Louis, P., Zhang, J., 2014. Noisy introspection in the 11–20 game. The Economic Journal.
Goldstein, I., Pauzner, A., 2004. Contagion of self-fulfilling financial crises due to diversification of investment portfolios. Journal of Economic Theory 119, 151–183.
Grosskopf, B., 2003. Reinforcement and directional learning in the ultimatum game with responder competition. Experimental Economics 6, 141.
Grosskopf, B., Nagel, R., 2008. The two-person beauty contest. Games and Economic Behavior 62, 93–99.
Grueneisen, S., Wyman, E., Tomasello, M., 2015. "I know you don't know I know ..." children use second-order false belief reasoning for peer coordination. Child Development 86 (1), 287–293.
Guererk, O., Irlenbusch, B., Rockenbach, B., 2006. The competitive advantage of sanctioning institutions. Science 312, 108–111.
Guesnerie, R., 1992. An exploration of the eductive justifications of the rational-expectations hypothesis. The American Economic Review, 1254–1278.
Güth, W., Kocher, M., Sutter, M., 2002. Experimental 'beauty contests' with homogeneous and heterogeneous players and with interior and boundary equilibria. Economics Letters 74, 219–228.
Güth, W., Schmittberger, R., Schwarz, B., 1982. An experimental analysis of ultimatum bargaining. Journal of Economic Behavior & Organization 3, 367–388.
Hanak, N., Sutan, A., Willinger, M., 2016. The Strategic Environment Effect in Beauty Contest Games. GREDEG Working Papers 2016-05. Groupe de REcherche en Droit, Economie, Gestion (GREDEG CNRS), University of Nice Sophia Antipolis.
Harbaugh, W.T., Krause, K., 2000. Children's altruism in public good and dictator experiments. Economic Inquiry 38 (1), 95–109.

Harrison, G.W., List, J.A., 2004. Field experiments. Journal of Economic Literature 42 (4), 1009–1055.
Haruvy, E., 2002. Identification and testing of modes in beliefs. Journal of Mathematical Psychology 46 (1), 88–109.
Haruvy, E., Noussair, C.N., Powell, O., 2014. The impact of asset repurchases and issues in an experimental market. Review of Finance 18 (2), 681–713.
He, Z., Xiong, W., 2012. Dynamic debt runs. The Review of Financial Studies 25, 1799–1843.
Heemeijer, P., Hommes, C., Sonnemans, J., Tuinstra, J., 2009. Price stability and volatility in markets with positive and negative expectations feedback: an experimental investigation. Journal of Economic Dynamics and Control 33.
Heinemann, F., Nagel, R., Ockenfels, P., 2004. The theory of global games on test: experimental analysis of coordination games with public and private information. Econometrica 72, 1583–1599.
Heinemann, F., Nagel, R., Ockenfels, P., 2009. Measuring strategic uncertainty in coordination games. The Review of Economic Studies 76, 181–221.
Heinemann, F., Noussair, C., 2015. Macroeconomic experiments. Journal of Economic Studies 42 (6), 930–942.
Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., 2001. In search of homo economicus: behavioral experiments in 15 small-scale societies. The American Economic Review 91, 73.
Herrmann, B., Thöni, C., Gächter, S., 2008. Antisocial punishment across societies. Science 319 (5868), 1362–1367.
Ho, T.H., Camerer, C., Weigelt, K., 1998. Iterated dominance and iterated best response in experimental "p-beauty contests". The American Economic Review 88 (4), 947–969.
Holden, S., Natvik, G., Vigier, A., 2014. An equilibrium theory of credit rating. Mimeo. University of Oslo.
Holt, C.A., 1995. Industrial organization: a survey of laboratory research. In: Handbook of Experimental Economics, vol. 349, pp. 402–403.
Holt, C.A., Laury, S.K., 2002. Risk aversion and incentive effects. The American Economic Review 92 (5), 1644–1655.
Hommes, C., 2011. The heterogeneous expectations hypothesis: some evidence from the lab. Journal of Economic Dynamics and Control 35 (1), 1–24.
Hommes, C., 2013. Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems. Cambridge University Press, Cambridge.
Hommes, C., Sonnemans, J., Tuinstra, J., Van de Velden, H., 2005. Coordination of expectations in asset pricing experiments. The Review of Financial Studies 18 (3), 955–980.
Hommes, C., Sonnemans, J., Tuinstra, J., Van De Velden, H., 2007. Learning in cobweb experiments. Macroeconomic Dynamics 11 (S1), 8–33.
Hommes, C., Sonnemans, J., Tuinstra, J., Van de Velden, H., 2008. Expectations and bubbles in asset pricing experiments. Journal of Economic Behavior & Organization 67 (1), 116–133.
Hossain, T., Morgan, J., 2006. Plus shipping and handling: revenue (non) equivalence in field experiments on eBay. Advances in Economic Analysis & Policy 6 (2), 1429.
Kagel, J., 1995. Auctions: a survey of experimental research. In: Kagel, John H., Roth, Alvin E. (Eds.), Handbook of Experimental Economics.
Kagel, J.H., Levin, D., 1986. The winner's curse and public information in common value auctions. The American Economic Review, 894–920.
Kagel, J., Levin, D., 1993. Independent private value auctions: Bidder behaviour in first-, second-and third-price auctions with varying numbers of bidders. The Economic Journal 103, 868–879.
Kagel, J., Levin, D., 2002. Common Value Auctions and the Winner's Curse. Princeton University Press, Princeton, NJ.
Kagel, J., Levin, D., 2016. Auctions: a survey of experimental research. In: Kagel, J., Roth, A. (Eds.), Handbook of Experimental Economics, vol. 2.
Kagel, J., Roth, A. (Eds.), 1995. The Handbook of Experimental Economics. Princeton University Press, Princeton, NJ.

Kagel, J., Roth, A. (Eds.), 2016. The Handbook of Experimental Economics, vol. 2. Princeton University Press, Princeton, NJ.

Kahneman, D., 2011. Thinking, Fast and Slow. Macmillan.

Kahneman, D., Tversky, A., 1979. Prospect theory: an analysis of decision under risk. Econometrica 47, 263–292.

Kaldor, N., 1934. The equilibrium of the firm. The Economic Journal 47, 60–76.

Karp, N., Lee, I.H., Mason, R., 2007. A global game with strategic substitutes and complements. Games and Economic Behavior 60, 155–175.

Keser, C., 1996. Voluntary contributions to a public good when partial contribution is a dominant strategy. Economics Letters 50 (3), 359–366.

Ketcham, J., Smith, V.L., Williams, A.W., 1984. A comparison of posted-offer and double-auction pricing institutions. The Review of Economic Studies 51, 595–614.

Kets, W., 2012. Learning with fixed rules: the minority game. Journal of Economic Surveys 26 (5), 865–878.

Kets, W., Sandroni, A., 2017. A Theory of Strategic Uncertainty and Cultural Diversity. Working paper.

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

Keynes, J.M., 1973. In: Moggeridge, D. (Ed.), The Collected Writings of J.M. Keynes. XIV. Macmillan for the Royal Economic Society, London, pp. 492–493.

Khaw, M.W., Stevens, L., Woodford, M., 2017. Discrete adjustment to a changing environment: experimental evidence. Journal of Monetary Economics 91, 88–103.

Kirman, A., 1992. Whom or what does the representative individual represent? The Journal of Economic Perspectives 6, 117–136.

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

Koszegi, B., Rabin, M., 2006. A model of reference-dependent preferences. The Quarterly Journal of Economics 121, 1133–1165.

Lagunoff, R., Matsui, A., 1997. Asynchronous choice in repeated coordination games. Econometrica 65 (6), 1467–1477.

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, A., 1981. Concours résultats complets. Les victimes se sont plua jouer. Jeux et Stratégie, 10–11.

Ledyard, J.O., 1995. Public goods: a survey of experimental research. In: Kagel, J., Roth, A. (Eds.), Handbook of Experimental Economics, pp. 111–194.

Lei, V., Noussair, C.N., 2002. An experimental test of an optimal growth model. The American Economic Review 92 (3), 549–570.

Levine, S.S., Bernard, M., Nagel, R., 2017. Strategic intelligence: the cognitive capability to anticipate competitor behavior. Strategic Management Journal 38 (12), 2390–2423.

Marcet, A., Sargent, T.J., 1989. Convergence of least squares learning mechanisms in self-referential linear stochastic models. Journal of Economic Theory 48 (2), 337–368.

Maskin, E., Tirole, J., 1987. Correlated equilibria and sunspots. Journal of Economic Theory 43, 364–373.

Matsui, A., Matsuyama, K., 1995. An approach to equilibrium selection. Journal of Economic Theory 65 (2), 415–434.

Matsuyama, K., 1991. Increasing returns, industrialization, and indeterminacy of equilibrium. The Quarterly Journal of Economics 106 (2), 617–650.

Matsuyama, K., 1992. Custom Versus Fashion: Path-Dependence and Limit Cycles in a Random Matching Game. Discussion Paper 1030. Department of Economics, Northwestern University.

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

Mauersberger, F., 2018. Thompson Sampling: Endogenously Random Behavior in Games and Markets. SSRN Working Paper.

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., 1990. Rationalizability, learning and equilibrium in games with strategic complementarities. Econometrica 58, 1255–1277.

Milgrom, P., Roberts, J., 1991. Adaptive and sophisticated learning in normal form games. Games and Economic Behavior 3, 82–100.

Mitzkewitz, M., Nagel, R., 1993. Experimental results on ultimatum games with incomplete information. International Journal of Game Theory 22, 171–198.

Monaco, A.J., Sabarwal, T., 2016. Games with strategic complements and substitutes. Economic Theory 62 (1–2), 65–91.

Morris, S., 2014. Coordination, timing and common knowledge. Research in Economics 68 (4), 306–314.

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

Morris, S., Shin, H.S., 2001. Rethinking multiple equilibria in macroeconomic modeling. In: NBER Macroeconomics Annual 2000, vol. 15. MIT Press, pp. 139–182.

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

Moulin, H., 1986. Game Theory for the Social Sciences. NYU 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.

Nagel, R., 1995. Unraveling in guessing games: an experimental study. The American Economic Review 85, 1313–1326.

Nagel, R., 2018. The rational man as a tuning fork. https://www.upf.edu/documents/8394861/161895668/a+change+-+the+rational+man+as+a+tuning+fork.pdf/81440ce1-967e-82f6-3acd-5ba1bd287ef7.

Nagel, R., Bayona, A., Kheirandish, R., Mousavi, S., 2016. Reinhard Selten, the dualist. In: Frantz, R., Chen, S.H., Dopfer, K., Heukelom, F., Mousavi, S. (Eds.), Routledge Handbook of Behavioral Economics.

Nagel, R., Bosch-Domenech, A., Charness, G., Garcia-Montalvo, J., Hurkens, S., Lopez Nicholas, A., Matthies, N., Rockenbach, B., Satorra, A., Selten, R., Thaler, R.H., 2018. Dataset One, Two, (Three), Infinity, ...: Newspaper and Lab Beauty-Contest Experiments. https://x-econ.org/xecon/#!Detail/10.23663/x2544.

Nagel, R., Bühren, C., Frank, B., 2017. Inspired and inspiring: Hervé Moulin and the discovery of the beauty-contest game. Mathematical Social Sciences 90, 191–207.

Nagel, R., Vriend, N., 1999. An experimental study of adaptive behavior in an oligopolistic market game. Journal of Evolutionary Economics 9, 27–65.

Nash, J., 1951. Non-cooperative games. Annals of Mathematics, 286–295.

Nerlove, M., 1958. Adaptive expectations and cobweb phenomena. The Quarterly Journal of Economics 72, 227–240.

Niederle, M., 2016. Gender. In: Kagel, J., Roth, A. (Eds.), Handbook of Experimental Economics, vol. 2.

Niederle, M., Segal, C., Vesterlund, L., 2013. How costly is diversity? Affirmative action in light of gender differences in competitiveness. Management Science 59, 1–16.

Obstfeld, M., 1996. Models of currency with self-fulfilling features. European Economic Review 40, 1037–1047.

Ochs, J., 1995. Coordination problems. In: Kagel, John H., Roth, Alvin E. (Eds.), Handbook of Experimental Economics, pp. 195–251.

Orland, A., Roos, M.W., 2013. The New Keynesian Phillips curve with myopic agents. Journal of Economic Dynamics and Control 37 (11), 2270–2286.

Östling, R., Wang, J.T., Chou, E.Y., Camerer, C.F., 2011. Testing game theory in the field: Swedish LUPI lottery games. American Economic Journal: Microeconomics 3, 1–33.

Oxford English Dictionary Online, 2018. Oxford University Press. http://www.oed.com. (Accessed 23 March 2018).

Palacios-Huerta, I., 2003. Professionals play minimax. The Review of Economic Studies 70 (2), 395–415.

Palfrey, T.R., 2016. Experiments in political economy. In: Kagel, J., Roth, A. (Eds.), Handbook of Experimental Economics, vol. 2, pp. 347–434.

Pfajfar, D., Zakelj, B., 2014. Experimental evidence on expectation formation. Journal of Economic Dynamics and Control 44, 147–168.
Plott, C.R., Smith, V.L., 1978. An experimental examination of two exchange institutions. The Review of Economic Studies 45 (1), 133–153.
Poe, E.A., 1958. The purloined letter. In: The Complete Tales and Poems of Edgar Allen Poe. Random House, NY, pp. 215–216.
Raafat, R.M., Chater, N., Frith, C., 2009. Herding in humans. Trends in Cognitive Sciences 13, 420–428.
Rabin, M., 1993. Incorporating fairness into game theory and economics. The American Economic Review 83, 1281–1302.
Rabin, M., 2002. Inference by believers in the law of small numbers. The Quarterly Journal of Economics 117, 775–816.
Rapoport, A., Guyer, M., Gordon, D., 1976. The 2x2 Game. The University of Michigan Press, Ann Arbor.
Rochet, J.-C., Vives, X., 2004. Coordination failures and the lender of last resort: was Bagehot right after all? Journal of the European Economic Association 2, 1116–1147.
Rogers, B.W., Palfrey, T.R., Camerer, C.F., 2009. Heterogeneous quantal response equilibrium and cognitive hierarchies. Journal of Economic Theory 144 (4), 1440–1467.
Roos, M.W., Luhan, W.J., 2013. Information, learning and expectations in an experimental model economy. Economica 80 (319), 513–531.
Rose, T., 2016. The End of Average: How to Succeed in a World That Values Sameness. Penguin, UK.
Roth, A., 2016. Experiments in market design. In: Kagel, J., Roth, A. (Eds.), Handbook of Experimental Economics, vol. 2, pp. 290–346.
Roth, A., Erev, I., 1995. Learning in extensive-form games: experimental data and simple dynamic models in the intermediate term. Games and Economic Behavior 8, 164–212.
Roth, A., Marilda, A., Sotomayor, O., 1990. Two-Sided Matching. A Study in Game-Theoretic Modeling and Analysis. Cambridge University Press.
Roth, A., Prasnikar, V., Okuno-Fujiwara, M., Zamir, S., 1991. Bargaining and market behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: an experimental study. The American Economic Review 81, 1068.
Rubinstein, A., 1982. Perfect equilibrium in a bargaining model. Econometrica, 97–109.
Rubinstein, A., 2007. Instinctive and cognitive reasoning: a study of response times. The Economic Journal 117 (523), 1243–1259.
Rubinstein, A., Tversky, A., 1993. Naive Strategies in Zero-Sum Games. Working Paper 17-93. The Sackler Institute of Economic Studies.
Rubinstein, A., Tversky, A., Heller, D., 1996. Naïve strategies in competitive games. In: Albers, W., Güth, W., Hammerstein, P., Moldovanu, B., van Damme, E. (Eds.), Understanding Strategic Interaction – Essays in Honor of Reinhard Selten. Springer-Verlag, Berlin, pp. 394–402.
Sargent, T.J., 1993. Bounded Rationality in Macroeconomics. Clarendon Press, Oxford.
Sauermann, H., Selten, R., 1959. Ein Oligopolexperiment. Zeitschrift für die Gesamte Staatswissenschaft (Journal of Institutional and Theoretical Economics) 3, 427–471.
Sbriglia, P., 2008. Revealing the depth of reasoning in p-beauty contest games. Experimental Economics 11 (2), 107–121.
Schaal, E., Taschereau-Dumouchel, M., 2015. Coordinating business cycles. Mimeo. NYU.
Schelling, Thomas, 1960. The Strategy of Conflict. Harvard University Press, Cambridge, MA.
Selten, R., 1965. Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit: Bestimmung des dynamischen Preisgleichgewichts. Zeitschrift für die Gesamte Staatswissenschaft (Journal of Institutional and Theoretical Economics) 121, 301–324.
Selten, R., 1967. Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments. In: Sauermann, H. (Ed.), Beiträge zur experimentellen Wirtschaftsforschung. Mohr, Tübingen, pp. 136–168.
Selten, R., 1975. Reexamination of the perfectness concept for equilibrium points in extensive games. International Journal of Game Theory 4 (1), 25–55.
Selten, R., 1991. Anticipatory learning in two-person games. In: Selten, Reinhard (Ed.), Game Equilibrium Models I: Evolution and Game Dynamics. Springer, New York, Berlin, London, Tokyo, pp. 98–154.

Selten, R., Mitzkewitz, M., Uhlich, G.R., 1997. Duopoly strategies programmed by experienced players. Econometrica, 517–555.

Selten, R., Pittnauer, S., Hoehnisch, M., 2012. Dealing with dynamic decision problems when knowledge of the environment is limited: an approach based on goal systems. Journal of Behavioral Decision Making 25, 443–457.

Selten, R., Stoecker, R., 1986. End behavior in sequences of finite Prisoner's Dilemma supergames: a learning theory approach. Journal of Economic Behavior & Organization 7 (1), 47–70.

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., 1996. The Sciences of the Artificial. MIT Press.

Sims, C.A., 2003. Implications of rational inattention. Journal of Monetary Economics 50 (3), 665–690.

Slonim, R.L., 2005. Competing against experienced and inexperienced players. Experimental Economics, 55–75.

Smith, A., Lohrenz, T., King, J., Montague, P.R., Camerer, C.F., 2014. Irrational exuberance and neural crash warning signals during endogenous experimental market bubbles. Proceedings of the National Academy of Sciences 111 (29), 10503–10508.

Smith, V.L., 1962. An experimental study of competitive market behavior. Journal of Political Economy 70 (2), 111–137.

Smith, V.L., Suchanek, G.L., Williams, A.W., 1988. Bubbles, crashes, and endogenous expectations in experimental spot asset markets. Econometrica 56, 1119–1151.

Sonnemans, J., Tuinstra, J., 2010. Positive expectations feedback experiments and number guessing games as models of financial markets. Journal of Economic Psychology 31, 964–984.

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., Wilson, P.W., 1994. Experimental evidence on players' models of other players. Journal of Economic Behavior & 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, 218–254.

Steele, C.M., 1997. A threat in the air: how stereotypes shape intellectual identity and performance. The American Psychologist 52 (6), 613–629.

Strzalecki, Tomasz, 2014. Depth of reasoning and higher-order beliefs. Journal of Economic Behavior & Organization, 108–122.

Sunder, S., 1995. Experimental asset markets: a survey. In: Kagel, John H., Roth, Alvin E. (Eds.), Handbook of Experimental Economics.

Sunder, S., 2007. What have we learned from experimental finance? In: Developments on Experimental Economics. Springer, Berlin, Heidelberg, pp. 91–100.

Sutan, A., Willinger, M., 2009. Guessing with negative feedback: an experiment. Journal of Economic Dynamics and Control 33 (5), 1123–1133.

Szkup, M., Trevino, I., 2011. Costly Information Acquisition in a Speculative Attack: Theory and Experiments. Working paper.

Taylor, S.E., Thompson, S.C., 1982. Stalking the elusive vividness effect. Psychological Review 89 (2), 155–181.

Topkis, D., 1979. Equilibrium points in non-zero sum n-person submodular games. SIAM Journal on Control and Optimization 17 (6), 773–787.

Tverksy, A., Kahneman, D., 1991. Loss aversion in riskless choice: a reference-dependent model. The Quarterly Journal of Economics 106, 1039–1061.

Tversky, A., Kahneman, D., 1974. Judgment under uncertainty: heuristics and biases. Science 185, 1124–1131.

Van Huyck, J., Battalio, R., Beil, R., 1990. Tacit coordination games, strategic uncertainty, and coordination failure. The American Economic Review 80, 234.

Van Huyck, J., Battalio, R., Beil, R., 1991. Strategic uncertainty, equilibrium selection, and coordination failure in average opinion games. The Quarterly Journal of Economics 106, 885–910.

Vesterlund, L., 2016. Using experimental methods to understand why and how we give to charity. In: Kagel, J., Roth, A. (Eds.), Handbook of Experimental Economics, vol. 2.

Vives, X., 1990. Nash equilibrium with strategic complementarities. Journal of Mathematical Economics 19, 305–321.

Vives, X., 2010. Information and Learning in Markets: The Impact of Market Microstructure. Princeton University Press, Princeton and Oxford.

Walker, M., Wooders, J., 2001. Minimax play at Wimbledon. The American Economic Review 91 (5), 1521–1538.

Walsh, C.E., 2010. Monetary Theory and Policy, 3rd edition. MIT Press, Cambridge.

Weber, R., Camerer, C.F., Rottenstreich, Y., Knez, M., 2001. The illusion of leadership in weak-link coordination games. Organizational Science 12, 582–598.

Weitzman, M.L., 1982. Increasing returns and the foundations of unemployment theory. The Economic Journal 92, 787–804.

Woodford, M., 1990. Learning to believe in sunspots. Econometrica 58, 277–307.

Woodford, M., 2003. Interest and Prices: Foundations of a Theory of Monetary Policy. Princeton University Press.

Woodford, M., 2018. Monetary Policy Analysis when Planning Horizons Are Finite. Working paper prepared for the NBER Macroeconomics Annual Conference.

Zermelo, E., 1904. Beweis, dass jede Menge wohlgeordnet werden kann (reprint). Mathematische Annalen 59, 514–516.