<https://helda.helsinki.fi>

Experimentation in economics

Nagatsu, Michiru

Routledge

2021-11-30

Nagatsu , M 2021 , Experimentation in economics . in C Heilmann & J Reiss (eds) ,
Routledge Handbook of Philosophy of Economics . Routledge , New York , pp. 329-342 .

<http://hdl.handle.net/10138/339402>

submittedVersion

Downloaded from Helda, University of Helsinki institutional repository.

This is an electronic reprint of the original article.

This reprint may differ from the original in pagination and typographic detail.

Please cite the original version.

ARTICLE

Experimentation in economics

Michiru Nagatsu^a

^aPractical philosophy and Helsinki Institute of Sustainability Science, University of Helsinki, Finland

ARTICLE HISTORY

Compiled January 26, 2021

ABSTRACT

This chapter provides an overview of different economics experiments, with a particular focus on different epistemic and non-epistemic interests underlying them. The chapter also discusses two types of philosophical interests in experimentation in economics, a general interest in the logic of inductive inferences and a more fine-grained interest in the divergence and convergence between economic and psychological perspectives in experimental practices in economics. Several experimental paradigms are used for illustrations, including public goods games, bargaining games, double auctions with the induced value method and random assignment method, and market entry games.

KEYWORDS

Experimental economics; experimental games; behavioral economics; co-production; severe testing; heterogeneity; psychology and economics; behavioral econometrics

JEL CLASSIFICATION

B29; B4; C7; C9; D91

1. Introduction

Why is experimentation in economics an interesting topic for philosophers? Güth and Kliemt (2017) note that “[t]o what extent the philosopher regards the toolbox of experimental economics as well as the results created by its use as philosophically relevant depends obviously on the fundamental conception of philosophy she endorses.” In this chapter, however, I shall not start from any fundamental conception of philosophy. Nor shall I argue for any particular one drawing on experimental economics. Instead, I will describe two main ways in which the methodology of experimental economics has been interesting philosophically, namely as an embodiment of the logic of inductive inferences and as a field in which interdisciplinary relations between economics and psychology become salient. As an illustration of how economics experiments can be conducted to illuminate philosophical topics such as the nature of signalling, language, and communication, see Rubin et al. (2019).

This article is prepared for Julian Reiss and Conrad Heilmann (eds.) *Routledge Handbook of Philosophy of Economics*.

CONTACT Michiru Nagatsu. Email: michiru.nagatsu@helsinki.fi

I will start by introducing three types of typologies of economics experiments in the next section. The typologies are meant to give the reader an overview of experimentation in economics, but they are also intended for revealing different epistemic and practical interests of experimental economists. In Sections 3 and 4, I turn to the philosophical relevance of the methodology of experimental economics. Section 3 briefly outlines how a general normative theory of inductive inferences, more specifically the severe testing approach, can be abstracted from the practice. Section 4 zooms in and discusses discrepancies in practices, often associated with the differences between economics and psychology, and how to reconcile these differences. Section 5 concludes. As illustrations I will draw on several experimental paradigms, such as a public goods game, a double auction, bargaining games, endowment effect experiments, and a market entry game. Sufficient details of these experiments are provided, but those who are interested in them should consult textbooks e.g. Camerer et al. (2003); Bardsley et al. (2010); Dharm (2016); Jacquemet and l'Haridon (2018) for more details.

2. Overview: typologies of experiments in economics

Before proceeding to the discussion of philosophical interests in experimentation in economics, it will be useful to provide a few typologies for distinguishing different kinds of experiments in economics. Since experiments are not a natural kind, there is no single correct way to do this. I will introduce three typologies in each subsections, coming from the philosophy of science, experimental economics, and science and technology studies. These typologies reveal different epistemic and non-epistemic interests underlying experiments in economics.

2.1. *The standard typology*

Economic experiments primarily concern economic agents' incentivized—motivated and consequential—choices and their aggregate outcomes. This central focus on incentivized choices makes them distinct from other types of experiments involving human subjects, such as biomedical, psychological and sociological experiments. One natural way to categorize economic experiments is thus by the type of the contexts in which those choices are made. It is customary to categorize economic experiments as three types, decision-theoretic, game-theoretic and market experiments, which corresponds to distinct choice situations and theoretical frameworks used to model them.² Table 1 summarizes these types in columns.

Decision-theoretic experiments concern individual choices that affect own pecuniary outcomes, while game-theoretic and market experiments concern choices that affect own as well as others' outcomes. An imperfect but intuitive way to distinguish these experiments is to look at the number of participants required: a decision-theoretic experiment can be conducted in principle with one participant, exposing her to a set of risky or intertemporal decision problems. A larger subject is required for reasons related to statistical significance and representativeness. Independently from those statistical reasons, a game-theoretic experiment requires at least two participants whose decisions are consequential to each others' pecuniary outcomes.³ A market experiment

²This way is also apparent in Roth (1993), which suggests that the three-part typology partly corresponds to the history of experimental economics.

³The dictator game (see later) is an exception in the sense that one of the two players does not make any choice. But for the experiment to count as game-theoretic, the presence of that passive player is needed.

Table 1. The three-part typology of experiments in economics

dimension\type	decision-theoretic	game-theoretic	market
1 theoretical models	expected utility theory, discounted utility model	non-cooperative games (Nash play; selfishness)	matching and auction theory (e.g. Vickrey auction)
2 measured constructs	risk preferences, time preferences, subjective probability	social preferences, strategic reasoning, expectations	market efficiency, robustness, strategy-proofness
3 well-established phenomena (‘exhibits’)	preference reversals, Allais’s paradox	‘fair’ offers in the ultimatum game, decline of contributions in the public goods game	convergence on equilibrium prices in Smith’s double auction
4 traditions	<div>testers</div> <div>builders</div>		

requires more than two participants, although how many are required depends on the market that is of interest. For example, if the experiment is about a duopoly market, then the participants can be as few as two.

In terms of the theoretical models underlying the experiments (the first row in Table 1) decision-theoretic experiments are based on models of individual choice, such as expected utility theory and the discounted utility model. The distinction between game-theoretic and market experiments is not so clear in this sense, since game theory (both non-cooperative and cooperative variants) is also essential for theoretical modelling of market experiments. The distinction becomes clearer, however, once one considers the target constructs of measurement (the second row of Table 1). In game-theoretic experiments, the experimenter’s interests are typically individual attributes such as preferences and strategic rationality: Do participants care only about own pecuniary gains and losses, or do they also care about those of others? What is their depth of strategic reasoning? In their focus on individual properties, game-theoretic experiments are similar to decision-theoretic ones. In market experiments, in contrast, the focus is on the system-level properties such as the efficiency of an experimental market with respect to the competitive prices, or the strategy-proofness of certain auction rules: Do those rules encourage participants to state their true valuation of goods, or do they incentivize them to strategically deviate from them? These are informational and allocational properties of an experimental system as a whole, consisting of human agents and a set of rules defining their actions and interactions as well as the information available to them.

Guala’s (2007) distinction between “the tester’s tradition” and “the builder’s tradition” roughly corresponds to the focus on system properties and the focus on individual properties, respectively, in experimental economics (the fourth row in Table 1). I will follow this terminology after making one qualification: not all experiments on individual properties are about testing theoretical hypotheses. Bardsley et al. (2010) argue that a lot of experimental economists’ effort goes to establishing phenomena—what Daniel Kahneman once called “the bottled phenomena”—and analyzing their behavior. Sugden (2005) uses the term an *exhibit* to refer to an experimental design that reliably produces an interesting phenomenon. An important function of this terminology is to make a distinction between two related but distinct practices, namely producing

a phenomenon and explaining it. Experimenters may disagree over what explains a particular phenomenon, but they disagree less on the fact that such a phenomenon can be produced reliably. In this theory-independent sense, the decision-theoretic experiments that demonstrate Allais' paradox⁴ and Vernon Smith's double auction experiments that demonstrate the convergence on competitive equilibrium prices and quantities of traded goods (see the third row of Table 1) are equally exhibits, although the former seems to refute the received theory (expected utility theory), while the latter seems to confirm it (the competitive market hypothesis). With this proviso in mind, the distinction between testers and builders are useful in contrasting two distinct loci of epistemic interests in economic experiments, namely the individual and system properties. Of course, since the former is part of the building blocks of any economic system, the builders have to pay attention to the individual properties to the extent that they matter to the design of the system. This concern is most salient in the public goods game, in which the two traditions intersect.

A public goods game is an experiment in which multiple players ($n \geq 2$) independently choose how to allocate the experimentally endowed money into private and public investments, which yield the payoff (π_i) defined as follows:

$$\pi_i = e_i - c_i + \frac{a \sum_k c_k}{N} \quad (1)$$

where $N > a > 1$, or equivalently $1 > \frac{a}{N} > \frac{1}{N}$. In words, each individual (i) is endowed with a certain amount of money (e_i), a fraction of which (c_i) she can contribute to the public investment ($e_i \geq c_i \geq 0$). Everyone's contribution is aggregated ($\sum_k c_k$), and then multiplied by a factor a , and finally equally divided by all the participants (N), regardless of their individual contributions. The condition $a > 1$ makes the public investment profitable as a collective endeavor, while the condition $N > a$ makes it always individually more profitable if one contributes nothing. Notice that equation (1) represents the individual perspective: e_i stays in your pocket as it is—by factor 1—whereas the contribution benefits you only by the factor $\frac{a}{N} < 1$.⁵ For example, an experiment with 4 players with $a = 2$ satisfies these conditions. These conditions constitute what experimental economists call a *voluntary contribution mechanism*, or VCM.

Public goods games are an instructive experimental paradigm often used to illustrate how experiments in economics are conducted (e.g. see Guala 2009, 2012) for several reasons. For one, they capture the features of the ubiquitous social dilemmas in everyday life, and therefore helps the reader to grasp the point of experimentation.⁶ The most important one, I suspect, however, is that it is a hybrid experiment in the sense that the tester's and builder's traditions intersect. On the one hand, the game can be interpreted as a many-player, many-strategy variation of the well-known Prisoner's Dilemma game, which is represented as a two-by-two normal form game with two players with two strategies. On this view, the properties of individual players, such as their beliefs, preferences, and strategic rationality can be the focus of the experiment. In fact, from this perspective Camerer (2003, 46) criticizes Prisoner's Dilemma and

⁴Allais paradox is...see [chapter on decision theory] in this volume.

⁵ $\frac{a}{N}$ is called *marginal per capita return* or MPCr. Although it is usually represented simply as α , without the explicit connection to N , my description makes it explicit that the lower boundary is necessary to make the project collectively profitable.

⁶This does not mean, however, that experimental subjects immediately understand the structure of the game because experimental games are often presented in a very context-neutral fashion.

public goods games as ‘blunt tools for guiding theories of social preferences.’ On the other hand, the VCM can be interpreted as an institutional arrangement to provide a public good. To be sure, this is exactly the kind of institutional design that the standard theory tells economists not to rely on for the provision of public goods (instead it recommends involuntary measures such as taxation). However, as I mentioned, VCM captures the strategic situations underlying not only everyday social dilemmas but many economically important decentralized arrangements in the management of common resource pools, for instance. Because of its prevalence and centrality in economic and everyday lives, the builder’s tradition has taken the paradigm seriously, trying to understand its system-level properties: How should the group be formed? Should communication be allowed, and if so what information? How many rounds should the players play? Should the participants be allowed to punish free-riders at own cost? The ‘should’ in these questions indicate the underlying design goal of the builder’s, namely the more voluntary provision of public goods.

Public goods games can be a tricky example, however, because of their dual character. In this experimental paradigm, one of the key tasks for the builder who wishes to induce voluntary provision of public goods (without changing the game structure itself significantly, that is) is to understand and exploit the individual properties of the participants, in particular those that do not conform to the standard models of choice. For example, if a substantial portion of the participants are *conditional cooperators* who are willing to cooperate if sufficient others do, then one might be able to generate a self-fulfilling pattern of cooperation by making them expect that sufficient others will cooperate. If a portion of participants have the preference for the so-called strong reciprocity, then introducing a possibility to punish free-riders at own costs may deter the potential free-riders, even though such a costly punishment is but a second order public good.⁷ If the participants are so-called team reasoners, framing the game in one way or the other might encourage or discourage cooperation.

In contrast, in a typical market design experiment, social preferences and non-standard strategic reasoning are carefully suppressed. For example, consider Smith’s double auction. In this experiment, an equal number of buyers and sellers are given different pre-determined reservation prices of a unit of an experimental commodity, for buyers in terms of willingness to pay (WTP, highest payable price) and for sellers in terms of willingness to accept (WTA, lowest acceptable price). The buyers and sellers then negotiate bilaterally to make an agreement for trade, and this is repeated in several rounds. A buyer (she) and a seller (he) can make profit if they manage to trade the commodity lower than her WTP and higher than his WTA, whereas how much profit each makes depends on the agreed price. One of the important rules of this experimental market is that the trade that creates negative profit is not accepted (zero profit is acceptable). This rule effectively eliminates the possibility to altruistically benefit the other trading partner. Another important rule is that each participant knows only own WTP or WTA, so it is impossible to know the size of the collective surplus (the pie) from a potential trade (which is defined as the difference between her WTP and his WTA). So not only can a trader not make own choice altruistic from a subjective point of view, but also a pair of traders cannot establish an inter-subjective understanding of a particular trade agreement to be fair or not. Also, neither the pair nor the buyers or the sellers are given any team frame, nor collective consequences, so the participants do not have a salient team-frame. These rules effectively eliminates

⁷That is, if punishment, while individually costly, benefits others in the group, there is an incentive to free-ride on others’ costly policing effort.

motivations other than making as much profit as possible as an individual. It is not that the participants are selfish, but rather that the experiment is designed such that they behave as if they were, so that the experimenter does not have to rely on the complex motivations underlying interpersonal transactions to achieve design goals. In other auction experiments, too, the key design task is to make sure that revealing own true valuation of a good is the dominant strategy for all the participants, that is, best regardless of what others do. In this context game theory is used to detect and eliminate the sources of deviation from the true valuation.

In sum, although market experiments such as Smith’s double auction and public goods games share the interest in the design of efficient institutions, their routes to get there are rather different. The latter try to understand and exploit non-standard behavior for the common good (voluntary provision of public goods), whereas the former try to eliminate it for another type of common good (maximizing collective surplus from trades). This is why I characterize the latter as having both aspects of tester’s and builder’s traditions: in public goods games, the tester’s interests in understanding quarks of human behavior and builder’s interests in establishing robust systems converge. In fact, the methodological problems of public goods games Camerer (2003) highlighted from the tester’s perspective have been addressed by new designs such as the use of the strategy method and the incentivized elicitation of beliefs (Fischbacher et al. 2001), rather than by adopting other experimental paradigms that Camerer thinks sharper tools. This continuous use of public goods games in experimental economics suggests the importance of the paradigm for the both traditions.

In order to understand the difference between the two traditions, it is therefore instructive to look at another famous game-theoretic experiment, the ultimatum game. An ultimatum game is a simple two-person, two-move game with the following setup. Proposer is endowed with some amount of money, let’s say 10 euro, and proposes own share, $10 \geq x > 0$, to Receiver, who would receive the rest ($10 - x$). The Receiver has two choices, either to accept or reject the offer. If he accepts, then $\pi_P = x$ euro, and $\pi_R = 10 - x$ euro. If he rejects instead, then both gets nothing ($\pi_P = \pi_R = 0$). Notice that this game can be interpreted as zeroing in on the situation in which Smith’s trading pair was facing. Let’s assume that the seller’s WTA is 5 euro, and the buyer’s WTP 15 euro. The potential surplus from this trade is 10 euro, and any proposal from either side is a proposal to divide it in a particular way. After several offers and counter-offers, the final, take-it-or-leave-it proposal is made at some point. An ultimatum game artificially creates this last part of the bargaining process. As I said, however, Smith’s traders do not know the size of the pie because the other’s WTA or WTP is unknown. In contrast, the standard ultimatum game makes the size of the pie as common knowledge, making sharing a salient frame. Ultimatum games clearly belong to the tester’s tradition because the focus is on isolating and understanding the individual properties, in particular the participants’ other-regarding preferences. Although this is a game-theoretic experiment, the strategic aspect is kept minimum to focus on the preferences: since the game involves only two moves, the experimenter can assume common knowledge between both players about implications of their choices. Specifically, since Responder clearly sees that acceptance always gives him more profit than rejection ($10 - x > 0$), his rejection must be motivated by something other than material gain, e.g. anger at what he perceives as an unfair offer from Proposer. Proposer’s fair offer can also be attributed to either rational fear of rejection (anticipating Responder’s anger), or the other-regarding preference for a fair distribution of outcomes. To investigate which is more dominant, the game can be varied by removing Responder’s veto power (this variant is called the dictator game).

As this example of bargaining games (ultimatum and dictator variants) shows, the tester’s tradition is interested in revealing and measuring non-standard or behavioral characteristics of individuals, whereas the builder’s tradition tends to suppress them.

To sum up this subsection, the main messages of the standard typology are (i) that different theoretical models underlie different experiments; and (ii) that different epistemic interests in the individual and system-level properties can be distinguished. I will come back to the distinction between the tester’s and builder’s tradition later on in Section 4, but now that we have established the standard typology of economics experiments, it is time to see how it can be extended.

2.2. *From the lab to the field*

The recent rise of field experiments in economics and other social sciences motivates another type of typologies, defined along the dimension of laboratory vs. field experiments (e.g. Harrison and List 2004; Charness et al. 2013). Table 2 summarizes two representative typologies and how they relate to each other, in terms of four dimensions: experimental subject, reference commodity, location of experiment, and subject awareness of being part of an experiment, each of which takes two values.

Table 2. Typologies of experiments based on Lab-field-ness. Based on Favereau and Nagatsu (2020)

Typology by Charness et al. (2013)	lab	extra-lab		field	
Typology by Harrison & List (2004)	lab	artefactual field	framed field	n/a	natural field
1 Subject	students	nonstandard	nonstandard	nonstandard	nonstandard
2 Reference	imposed	imposed	field	field	field
3 Location/site	artefactual	artefactual	artefactual	natural	natural
4 Subject awareness	yes	yes	yes	yes	no

This type of typologies highlights a systematic way in which naturalism increases from lab to field experiments. On the left column is the archetype laboratory experiment: (1) participants are university students, who make choices to earn (2) imposed experimental currency (which is later converted to real local currency), (3) in a room with computers that connect different participants, (4) with full awareness that the experimenter is monitoring and recording their choices. The other extreme is a natural field experiment on the right, in which all these dimensions are modified. For example, the experimenter can study price discrimination by (1) automobile repairmen, who (2) choose an offer price to a potential customer (3) in their repair shops, (4) without knowing that the customer’s request for an offer is part of experimental interventions (Gneezy et al. 2012). In laboratory experiments, the setup itself is an important source of control for unknown variations (use of students with similar cognitive abilities, a training session to make sure that the students understand incentives, incentivizing them with the common currency, homogenizing their informational environment by having them sit in front of the same type of terminals, etc.). Field experiments, in contrast, crucially depends on the method of randomized assignment of treatments to the participants, in order to control indirectly for confounding factors. For example, half of the repairmen receive disabled customers, while the other half abled ones. Since this assignment is random, we can reason that the difference in the average price offers of these two groups are caused by the differences of treatments (abled vs. disabled customers asking for a quote), and not by other unknown factors that systematically differentiate the two groups (e.g. profit margins). In between these two extremes lie

hybrid experiments that have elements of both lab and field experiments. In principle one can bring any type of experiments (decision-theoretic, game-theoretic or market experiments) from the lab to the field, to different degrees. And these experiments can have a range of epistemic interests from comparative measurements of some individual properties (e.g. social preferences) to testing theoretical hypotheses. In the case of the natural field experiment with car repair shops, for example, the research goal was to inform theoretical debates between the taste-based discrimination hypothesis and the information-based discrimination hypothesis.

Although these typologies are useful in understanding how field experiments are related to laboratory ones, they do not show an important fact that the practices that economists call field experiments are not necessarily rooted in the tradition of laboratory experiments with theory-testing and establishing phenomena. Nagatsu and Favereau (2020) argue that field experiments in economics have in fact roots in two historical strands, the laboratory experimentation which wishes to systematically reduce the artificiality of laboratory setups (Lab-in-the-Field Experiments, LFEs), and the social experimentation with policy evaluation as the main goal (Randomized Field Experiments, RFEs). Philosophers discuss both strands in the context of the problem of *external validity*: How can we reliably infer that X causes Y in the target system from the knowledge that X causes Y in an experimental system?⁸ The LFEs' move from the lab to the field can be understood as an attempt to address this question by a systematic controlled variation of experimental setups themselves. In contrast, RFEs, in particular those in development economics, have gained prominence in the last two decades as the 'gold standard' to collect evidence on 'what works' as effective policy interventions, bypassing theoretical debates in development and behavioral economics. The RFE strand has encountered numerous methodological criticisms, many of which center around the problem of generalizability of average treatment effects, which mask the underlying data generating processes (e.g. subjects heterogeneity) and therefore cannot be easily extrapolated to other situations without a guide of theory (Deaton and Cartwright 2018). Since both LFEs and RFEs have different methodological limitations and affordances, they can be combined to address different issues of external validity in a complementary way (Favereau and Nagatsu 2020). In this chapter, I mainly discuss the LFEs [See Favereau's chapter in this volume for the RFEs strand.]

2.3. *From the lab to the wild*

Note that field experiments (in both LFEs and RFEs strands) are still characterized as economics experiments in the sense of aiming for controlled variations to achieve epistemic goals (measurement and causal inference). The notion of experimentation can be extended further to cover non-epistemic interests such as political persuasion and reform. From this perspective, which is more prominent in science and technology studies (STS) than in the philosophy of science, experiments are socio-technological devices to construct and interact with reality. STS scholars thus de-emphasize the lab-field divide, covering some activities 'in the wild' also as experimentation. Muniesa and Callon (2007) for example argue that economics experiments participate in the co-production of knowledge and governance, whether it is inside or outside of the laboratory wall. Table 3 summarizes the typology of Muniesa and Callon (2007) in column and their dimensions in row (row 4 shows their examples).

Superficially, Table 3 resembles Table 2: the field-ness seems to increase from left to

⁸For critical discussions of the external validity concept, see Jiménez-Buedo (2011); Reiss (2018).

Table 3. A threefold typology of economics experiments, based on Muniesa and Callon (2007)

dimension\typology	laboratory	platform	<i>in vivo</i>
1 location (materials)	lab closed	hybrid	open in the wild
2 manipulation (operations)	imposition of experimental protocols	simulation-by-modeling simulation-by-testing	injection of intervention into a black box
3 demonstration (acceptance)	academic publications =>limited	stakeholder engagement =>high, politicized	maximal stakeholder engagement
4 example	Elicitation of willingness to pay for GMO-free food products by Vickery auction	Optimization models for electricity market reform in France French consumerist magazines' product tests	Reform of stock exchange rules in the Paris Bourse

right as you move from laboratory experiments to *in vivo* experiments. Also the rows include the dimension of location, which also appears in Table 2. However, the typology of Muniesa and Callon (2007) covers a much wider range of experiments than the other. All experiments in Table 2 belong to laboratory experiments in this typology. And what they call platform experiments are not experiments in the standard sense, but more like public simulation with non-human subjects (e.g. commercial products and numbers in equations and computer programs). *In vivo* experiments are not experiments in the standard sense, either, since they are characterized as lacking control, not even a random allocation of the subjects into treatment and control groups. They are experiments in a more non-technical sense of involving trials and errors on the part of the experimenters and participants. As these points show, the typology is not interested in experiments as an epistemically prestigious method of inferences because of their high level of direct and indirect control of variables, as in the philosophical literature. Rather, the focus is on the unique ways in which economics experiments contribute to the so-called *co-production of knowledge and governance*.

Co-production in this context refers to two things. One concerns the plurality of subjects. The typology is conscious that the knowledge is produced not only by academic researchers qua experimenters (as in the laboratory experiments), but by other stakeholders including political and administrative decision makers, and those who will be affected by their decisions (citizens, companies, NGOs, etc.). As the number of stakeholders increases, the role of experiments as consensus-building and negotiation tools looms larger, as in the platform experiments. The example of the latter Muniesa and Callon (2007, 175) discuss is how ‘EDF (the state electricity company) succeeded in imposing its nuclear (and economic) model against the one defended by the CEA (the national agency for atomic energy)’ in France, through the engagement of stakeholders in econometric optimization models that simulated real-scale industrial experiences of market reforms.

The other meaning of co-production concerns the way in which knowledge produces a mode of governance as a by-product, or *vice versa*. This co-production character is

weakest in laboratory experiments, whose results are usually reported in academic and other types of publications to inform policy indirectly. It is important to note that the demonstration depends not only on the epistemic reliability of experimental results, but also on rhetorical qualities. For example, nudging policy based on behavioral economics is easy to understand because it mostly draws on accessible and intuitive individual psychology; RFEs are also easy to understand because of the simple way in which average treatment effects demonstrate ‘what works.’ However, although these publications can become influential and initiate a policy trend, their influence is still indirect. In contrast, in platform experiments, such as the simulation for the structural reform of the electricity market, the template of experiments is directly reflected in the political process of policy making. In this case, specifically, the simulation experiment created and analyzed different public investment strategies in terms of their costs and benefits, and eventually led to the political acceptance and administrative adoption of strategies deemed acceptable by the most stakeholders. Platform experiments not only inform policy but also *form* it by the very model-building practices (See e.g. Van Egmond and Zeiss 2010).

Co-production collapses, as it were, in *in vivo* experiments because the latter blur the boundary between the experimenter and the experimented, as in Muneisa and Callon’s (2007) example of the reform of the French stock market’s online architecture. In this case, both the actors (knowledge producer and user) and products (knowledge and practice) in experiments are merged and become inseparable.

Muneisa and Callon’s (2007) typology is tuned to understanding the builder’s tradition in experimental economics in a wider context. The typology highlights that this tradition is interested not only in building a desirable system in the lab, but in building one that will be part of the socio-technical systems that govern our economic lives. In fact, all these three types of experiments are utilized in both episodes that Guala (2005) uses to illustrate the builder’s tradition: the design of the bandwidths auction by the US Federal Communications Commission (FCC), and the design of the auction of oil leases in the Outer Continental Shelf (OCS) in the Gulf of Mexico. Worrapimphong et al. (2010) also detail how a variation of Voluntary Contribution Mechanism (VCM) of public goods games can be used in devising and implementing new fishery management strategies together with local fishermen and officials, mainly on platform experiments (see Redpath et al. 2018, for more examples of the constructive use of game theory).⁹ In these highly applied policy and management contexts, the experimenter does not stay in the laboratory, from which she reports experimental results and suggests their policy implications; she goes out in the wild, interacting with stakeholders and sometimes becoming one of the participants of the system she herself helped build.¹⁰

I summarize this section that has illustrated different typologies of economics experiments. Different typologies highlight some aspects of economics experiments and hide others. Table 1 shows basic categories of laboratory experiments, but I emphasize that the experimenter’s epistemic interests can diverge to studying individual and system-level properties of a particular experimental paradigm. Table 2 shows that experiments can have a degree of field-ness, but again I emphasize that the experimenters’ epis-

⁹Worrapimphong et al. (2010) also used non-incentivized role-playing games using fishermen as participants, and after that used a platform in which different harvesting strategies were simulated with researchers and fishermen. Although the researchers hoped to implement the agreed strategy *in vivo*, they could not continue the experiment further because mayor had changed and their experiments had lost political support.

¹⁰In the FCC auction case, some game theorists involved in the research were also hired as consultants in the actual auction.

temic interests can diverge, from hypothesis testing and the measurement of individual properties to policy evaluations. Table 3 relaxes the notion of experimentation associated with controlled variations in the philosophical literature, and highlights that the system-building tradition in experimental economics involves wider methods by which to involve stakeholders and co-produce experimental knowledge and governance practices. The last typology makes clear that when one studies experimenters' interests, one may need to pay attention not only to a range of epistemic interests but also practical and even political interests (See e.g. Mirowski and Nik-Khah 2007).

3. Philosophical interests: the logic of inductive inferences

In the last section, we have seen that the interests of economics experiments can range across different kinds of epistemic, practical and even political ones. But what are *the philosopher's interests* in studying these experiments? Understanding different interests underlying them is surely one, which motivated our overview of the typologies in the last section. In this and the next sections, I will describe in more detail two main methodological interests, continuing to add some of the experimental paradigms as illustrations. I will divide those interests in high-level and low-level and start with the former.

Perhaps one of the most prominent reasons why economics experiments interest philosophers of science is their methods to achieve reliable inductive inferences through systematic and controlled variations of experimental conditions. Guala (2012) thus sees experimental economics as a practice from which a general normative theory of inductive inferences can be abstracted. On this view, the experimental method is essentially 'an attempt to investigate causal influence of separate factors working in isolation.' (Guala 2012, 613) Specifically, when the experimenter varies only one factor as experimental and control conditions (denoted as X_e and X_c , or simply x and its absence), while holding all the other factors (denoted as K) constant, then the difference between the outcomes of the two conditions (denoted as $Y_e - Y_c$) can be reliably attributed to the difference between the two conditions, namely the experimental treatment x . Guala (2012) argues that this *model of the perfectly controlled experiment* captures the normative essence of various inferential strategies that experimental economists use.

Now consider a simple causal hypothesis H : x causes y (and $\neg H$ as it's falsity). We can think about a specific hypothesis that the Proposer's concern for fair distribution of income causes him to propose the Receiver the 50/50 divide of the pie in a bargaining game. The model of perfectly controlled experiment implies that in a good experiment, namely an experiment in which the experimenter controls for the background conditions (K) well, (a) the observed values of the variables X and Y are correlated if H (x causes y), and (b) they are not correlated if $\neg H$ (x does not cause y). Denote the confirming observation as E , and the disconfirming observation as $\neg E$. The experiment is a *severe test* of H if and only if the experimenter is likely to observe (a) E if H , and (b) $\neg E$ if $\neg H$. The test is severe in the sense that E supports H more than in the case in which E would be observed anyways had H been false. Is the ultimatum game a severe test of our H ? It is not because x —Proposer's concern for fair distribution of income—is not observable and not operationalized, that is, we do not observe variations of X . Specifically, there is another possibility that the fair offer is caused by the Proposer's fear of the Receiver's rejection that would lead him to earn nothing. In order to test this possibility, one can compare the original

game with a dictator game, in which the Receiver does not have any say in the game, and therefore the Proposer has no reason to fear his minimum offer to be rejected. Comparing the observations from the two games is a more severe test of the fairness hypothesis than its test in the ultimatum game because the former introduces the likelihood of observing $\neg E$ if $\neg H$. Of course, this controlled variation may introduce other uncontrolled confounding factors, compromising its severeness as a test of H (List 2007; Bardsley 2008). Nevertheless, a general lesson here is that severe testing guides an iterative process of experimentation, as well as the design of and inferences from single experiments.

Guala (2012) also emphasizes that a hypothesis that can be severely tested in an experiment is usually very specific and local (e.g. fair offers in bargaining games is robust and not due to a chance event), rather than high-level theoretical hypotheses or explanations (e.g. that fairness concern causes fair offers in bargaining situations; or that this game elicit the latent and general construct of ‘fairness’ of individuals). This is to say that the logic of severe testing guides not only theory-testing experiments but also exhibits experiments, namely the kind of experiments that mostly concerned with establishing robust phenomena. Similarly, severe testing also guides RFEs that philosophers have been criticizing for its inability to allow for the understanding of causal mechanisms underlying average treatment effects. The tested causal hypothesis in a RFE is fairly limited: the average treatment effect is caused by the treatment and nothing else. And testing this particular hypothesis severely does not eliminate controversies about many other hypotheses about why this causal relation holds.

In sum, on this view economics experiments are interesting because philosophers can extract an important normative principle of inductive inferences, such as severe testing. (for more elaboration of the severe test approach in contrast to other accounts of inductive inferences see Guala 2012, 3.3)

4. Philosophical interests: methodological divergence and convergence between economics and psychology

At certain level of abstraction, philosophical accounts of causal inferences make explicit common normative methodological principles that guide a wide range of experimental practices. This type of abstraction addresses one type of philosophical interests, but leaves another unsatisfied. That is: Can we find more specific normative standards against which we can evaluate experimenters’ disagreements? The typologies we reviewed in the last section suggests that the normative methodological standards may be different depending on one’s interests. But these typologies are not sufficient when the philosophers want to understand and reconcile controversies in which experimenters with apparently converging epistemic interests disagree over what should be the right account of a specific phenomenon, or the right approach to investigating that type of phenomena.

Controversies of this type are most salient in the debates in behavioral economics, whose central methods include experiments. Behavioral economics is a kind of inter-field between economics and psychology, and many practitioners have discussed differences between the two disciplines (e.g. Simon 1986; Kahneman 1988; Smith 1991; Lewin 1996; Rabin 1998, 2002; Hertwig and Ortmann 2001). Part of the core differences is, to use our terminology, that psychology is associated with the tester’s tradition, and economics with the builder’s. The former is interested in the individual-level properties of decision makers, mostly operating in decision-theoretic and game-theoretic

experiments, while the latter is interested in the aggregate-level properties of interactive systems, mostly dealing with game-theoretic and market experiments (see Table 1). However, there are some cases in which these interests converge, as I have illustrated with public goods experiments in Section 2. In this section, I will discuss two more cases, mostly drawing on Kahneman (1988). Both cases are market experiments, which Kahneman (1988) takes to jointly demonstrate the limits of the psychologist's contributions to economics. I update Kahneman's assessment, suggesting that there is room for further convergence between economics and psychology, drawing on the recent results from these experiments.

4.1. *Psychology explains the market*

The first case is a series of famous experiments on endowment effects (Kahneman 1988; Kahneman et al. 1990, 1991). Endowment effects refer to the phenomenon in which one's measure of willingness to accept (WTA) in a seller's position is higher than the measure of willingness to pay (WTP) in a buyer's position for the same good. In standard economic analysis, WTA and WTP are assumed to be approximately the same, such that a mere ownership does not confer any psychic benefit or reluctance to trade. Remember Smith's double auction, in which the experimenter assigns both buyers and sellers pre-determined WTP and WTA for a unit of the experimental commodity. This manipulation is called the *induced value technique*. Under this control, the WTP of a buyer who is told that a unit is worth 5 euro for him will likely be 5 euro, and so is the WTA of a seller who is told that a unit costs her 5 euro. This will not change whether the same participant is assigned the role of a seller or a buyer. Is this experiment a severe test of the hypothesis that $WTA = WTP$? It is not. Rather, the experimental design ensures that this equality holds. In order to test this hypothesis, another design is needed. Kahneman and colleagues devised such an experiment using a *random allocation technique*. In this design, the experimenter does not induce or impose values of a good to the participants, but instead assign them randomly into a seller and buyer role, by giving a random half of the participants goods such as coffee mugs and pens with a university logo (worth roughly five dollars). This way those who receive the good become potential sellers, and those who do not potential buyers. Then individuals from both groups can meet and negotiate bilaterally to agree on a trade, as in the Smith's double auction.¹¹

The key idea is that since this allocation is random, there is a 50/50 chance that a good is given to someone who values it more (and less) than the other participants. If $WTA = WTP$ for all individuals, then approximately half of the goods will be traded. We do not know the exact shapes of the supply and demand curves, unlike in the experiment with the induced value technique, but we can expect that they will be 'mirror images of each other' (Kahneman 1988, 15), leaving the volume of the trade approximately half of the number of the distributed goods. The result, however, is that the actual volume of the trade is about 50% of the prediction. As Kahneman (1988) notes, this is not a refutation of $WTA = WTP$ in all commodity markets; in fact, this equality holds up well in some situations, and not others, with unclear boundary between the

¹¹In the main design by Kahneman and colleagues, both groups fill in a questionnaire about their WTA and WTP, which enables the experimenters to construct the demand and supply curves and compute the market price of the mug at a particular round. The procedure is repeated several times, and a random draw determines which round will actually determine the participants' outcomes. Since they also conducted the negotiation variant with essentially the same results, I will focus on that variant in this presentation. For the original publication see Kahneman et al. (1990). For recent developments see Dharm (2016, 217-235).

two types of situations. However, these series of experiments demonstrate that investigations of potential individual-level psychological causes of the discrepancy—such as ‘loss aversion, or ambiguity in the value of the good in conjunction with anticipated regret’ (Kahneman 1988, 16)—can contribute to the understanding of the aggregate-level outcomes such as the volume of trades. The experiments on endowment effects are important for behavioral economics because it vindicates the relevance of psychological factors in economics: individual-level biases may bias system-level outcomes in a predictable way.

4.2. *Psychology does not explain the market*

Kahneman (1988) illustrates another market experiment that are in stark contrast to the experiments on endowment effects. He dubs it an N^* game, but the game is more commonly known as a market entry game. It is an n -player symmetric coordination game without communication: the market is profitable for entrants, but the marginal profit from entry decreases as the number of entrants increases, and beyond a certain market capacity (denoted by N^*) profit becomes negative. One can think of it as an n -person variation of the game of Chicken. In the original experiment, $N = 15$ and N^* was changed over a period of repetition within the range of $12 > N^* > 3$; the payoff to each person was \$.25 if one did not enter the market, and $$.25 + .5(N^* - E)$ if one did, where E is the number of total entrants.¹² If $E = N^*$ —if the number of actual entrants equals the capacity—both entrants and non-entrants receives the same \$.25 payoff. To Kahneman’s great surprise, E quickly converged to N^* in a few rounds, and stayed within the range $N^* + 1 \geq E \geq N^* - 1$ in the vast majority of trials. This is the implication of a mixed-strategy Nash equilibrium: each player decides to enter with a probability such that the expected payoff from entry equals that from non-entry—and is indifferent between entering and not entering in a steady state. The aggregate outcome from such individual strategies will result in $E \approx N^*$.¹³

What made Kahneman think that the result as ‘almost like magic’ is not that the equilibrium result was reached. He has observed that in a variation of the experiments on endowment effects. Rather, what was ‘bewildering’ was that ‘[t]he equilibrium outcome (which would be generated by the optimal policies of rational players) was produced in this case by a group of excited and confused people, who simply did not know what they were doing.’ (Kahneman 1988, p. 12) In other words, there is a causal gap between the aggregate phenomenon (the Nash equilibrium) and individual behavior (a mixture of non-Nash plays). The two lessons Kahneman learned from this experiment are (i) that the cognitive psychologist ‘has essentially nothing of interest to contribute’, and (ii) ‘that his bag of intellectual tools lack the powerful instrument of equilibrium explanations’ (Kahneman 1988, 13). The second lesson is trivially true if game theory is monopolized by economics, but the first one is more substantial, which I will examine.

In fact, the results of the N^* game experiment, conducted by Kahneman and the economists Richard Thaler and James Brander, has never been published in a peer-reviewed journal. This fact is consistent with Kahneman’s assessment that psychology has little to contribute to the understanding of the equilibrium results in market entry games. Alternatively, one can speculate that the reason why this was not published is

¹²We used the numbers as in Kahneman (1988), which are slightly different from those presented in his Nobel autobiography.

¹³See Rapoport and Seale (2008) for the overview of the market entry experiments.

because the message was not favoring their agenda to promote behavioral economics, and the priority was instead given to publishing results from endowment effects experiments (Kahneman et al. 1990), as well as ultimatum and dictator games (Kahneman et al. 1986a,b), both of which highlight the violation of the standard economics predictions while offering psychological explanations.¹⁴ In any case, later results from market entry games show that cognitive psychology *is* relevant to the understanding of market entry games. Amnon Rapoport and his colleagues have conducted a series of market entry games since 1995, and find (i) that when learning is allowed by design, individual behavior also converge on some individual-specific behavioral rules (for example conditionalizing one’s entrance decision on some cut-off market capacity c^*); and (ii) that the distribution of different cut-off rules is sufficient for generating the equilibrium results (Rapoport and Seale 2008). In particular, the study of (i) is informed by the theory of short-term memory and learning, to which psychologists have been contributing in behavioral game theory (see Camerer 2003, chapter 3). So it is not generally true that cognitive psychology has no potential contribution here. The relevant causal mechanisms underlying the equilibrium results concern both individual (learning of heuristics) and aggregate (combination of heterogeneous behavioral patterns), and the former is a psychological phenomenon. Moreover, some behavioral economists (e.g. Dharami 2016) contend that the type (ii) explanation is an alternative, more realistic explanation than the “as if” explanation by mixed-strategy Nash equilibrium.¹⁵ So the fact that psychologists do not have ‘the powerful instrument of equilibrium explanations’ might not even be the fundamental obstacle for psychologists to contribute to the study of the aggregate phenomenon.

The last point is most clearly demonstrated by recent developments in decision-theoretic experiments, in which the aggregate-level data are not explained by Nash equilibrium because they are not the results of interactive decision-making. However, this does not mean that the explanations of the results from these experiments are exhausted by individual-level psychological mechanisms. The outcome of a risky-choice experiments, for example, can be explained as resulting from either (i) the average behavior of one type of decision makers (e.g. following expected utility theory, prospect theory, and so on), or (ii) the combination of the behavior of these heterogeneous types of decision makers. The results of a decision-theoretic experiment can even be an outcome of the mixture of risky choice and inter-temporal choice, that is, caused by both risk and time preferences of any given individual decision maker (Andersen et al. 2008). When the experimenter tries to test the truth of one model of choice against another, (i) is implicitly assumed (model selection test). But the experimenter can also explicitly test this assumption, by making a structural model of the data-generating processes, assuming that multiple types of decision-makers exists, or that multiple types of decision-making mechanism operate inside one individual, or both (model specification test). As a statistical technique to do this, the maximum likelihood estimation of such a structural model with a mixture specification is used (Harrison 2019). This approach to explicitly addressing the heterogeneity of data-generating processes in experimental economics is called *behavioral econometrics* (Andersen et al. 2010), or *experimetrics* (Moffatt 2016). As it demonstrates, even in non-game theoretic

¹⁴James Blander points out one reason why the N* game was not completed was timing: Thaler’s visit to University of British Columbia in 1984-1985 ended, and Kahneman left the UBC the following academic year. But Blander also mentions that “we had not really found a good way to demonstrate the key insight we were trying to identify.” (Email correspondence with the author: 25 May 2020)

¹⁵Even Varian (2010), a popular textbook by an economist, is skeptical about the extent to which mixed-strategy Nash equilibrium explains the real aggregate phenomena in coordination.

and non-market experiments, explicitly modeling how individual-level choices give rise to aggregate-level data seems crucial for better understanding experimental results.

In particular, this approach can lead researchers to explain individual-level decision-making mechanisms in a way that goes beyond the standard utility maximization modeling framework. Andersen et al. (2014) for example illustrate how a non-maximization, dual-process model of individual risky choice can explain the experimental results using the behavioral econometrics approach. This is significant methodological progress from the psychological perspective, too, given that the continuous dependence on the “as if” models of utility maximization has been one of the main complaints about behavioral economics by some psychologists (e.g. Berg and Gigerenzer 2010).¹⁶

To summarize this section, Kahneman (1988) takes the endowment effect experiment to demonstrate one situation in which psychology can directly explain economic phenomena, while the N^* game to demonstrate another situation in which psychology has no contribution to make. Drawing on the recent experimental results and methodological developments in experimental economics, I re-evaluated this assessment, which was made more than 30 years ago, and also revisited by Kahneman in his Nobel Biography (Kahneman 2003). Put simply, my assessment is that Kahneman (1988) was too pessimistic about the psychologist’s contribution to economics. Moreover, this pessimistic assessment was based on the contingent dichotomy between psychology as responsible for the explanation of individual choices and economics of the aggregate choice data. This dichotomy has been challenged both empirically and methodologically.

5. Conclusion

In this chapter, I provided an overview of diverse experimental practices in economics, by way of reviewing three typologies of economics experiments (Section 2). These typologies highlighted that different experimental practices are motivated by different epistemic and non-epistemic interests, such as investigating individual and system-level properties of experimental systems, extrapolating experimental results to non-experimental setups, evaluating policy, constructing a new socio-technological system in the wild with stakeholders, etc. The philosopher’s interests in a normative theory of inductive inferences allows him to identify a general normative principle, or logic of inductive inferences, the severe testing approach, among others, that penetrates these different types of experimental practices (Section 3). At the same time, philosophical interests call for some explanation and evaluation of the existing controversies in experimental economics. I focused on a general question “Does economics need psychology?” to which different economists and psychologists would give different answers. To situate this question in the context of experimental economics, I examined Kahneman’s (1988) take on the bounds between economics and psychology that draws on two market experiments he has conducted, one famous and the other unknown in behavioral economics (Section 4). Drawing on the recent developments in these and other experiments, I argued for a kind of symmetrical convergence between

¹⁶Of course, this debate is confused with what different researchers mean by ‘realistic models.’ Concerning the realisticness of the structural equation models of the data-generating process, for example, Andersen et al. (2010, 554) state that the maximum likelihood approach is ‘an intellectual device to write out the process in as detailed a fashion as desired.’ Such a comment suggests that how realistic a model needs to be depends on the purposes of analysis, and that it is less fruitful to evaluate the realisticness of a given model independently from these purposes.

psychological and economic perspectives in experimental economics. Although behavioral economics has highlighted a particularly dramatic way in which psychology can be directly relevant to the study of economic phenomena (as in the endowment effect experiment), there are much more potentials for psychological and economic perspectives to inform each other in experimental economics. I have briefly mentioned the use of the theory of short-term memory and learning in behavioral game theory and the use of a non-optimization model of individual choice in decision-theoretic experiments.

I said *symmetrical* convergence because these developments point to the limit of framing the debate as one-sided as in the question “Does economics need psychology?” At least in some important class of economics experiments in which the interests in the individual- and aggregate-level phenomena converge, it seems that psychology also needs economics, as much as the latter needs the former.

Acknowledgements

I thank Francesco Guala, Judith Favereau, X, Y, Z for valuable comments on early drafts of this chapter.

Disclosure statement

The author is not aware of any potential conflict of interest in this research.

Funding

This research is conducted as part of the project: Model-building across disciplinary boundaries: Economics, Ecology, and Psychology (2016-2021), funded by the Academy of Finland (No. 294545).

Notes on contributor(s)

Michiru Nagatsu is an associate professor at the Helsinki Institute of Sustainability Science, and Practical Philosophy, the University of Helsinki. He runs *Economics and Philosophy Lab* and *HELSUS Methodology Lab*. His research uses a range of empirical approaches—including experimental philosophy, collaborations with scientists, interviews, integrated history and philosophy of science—to study conceptual and methodological questions in the philosophy of science.

References

- Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, 76(3):583–618.
- Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2010). Behavioral econometrics for psychologists. *Journal of Economic Psychology*, 31(4):553 – 576.
- Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2014). Dual criteria decisions. *Journal of Economic Psychology*, 41:101 – 113. From Dual Processes to Multiple Selves: Implications for Economic Behavior.

- Bardsley, N. (2008). Dictator game giving: altruism or artefact? *Experimental Economics*, 11:122–133.
- Bardsley, N., Cubitt, R., Loomes, G., Moffat, P., Starmer, C., and Sugden, R. (2010). *Experimental economics: Rethinking the rules*. Princeton University Press.
- Berg, N. and Gigerenzer, G. (2010). As-if behavioral economics: Neoclassical economics in disguise? *History of Economic Ideas*, 18(1):133–166.
- Camerer, C., Issacharoff, S., Loewenstein, G., O’Donoghue, T., and Rabin, M. (2003). Regulation for conservatives: Behavioral economics and the case for ‘asymmetric paternalism’. *University of Pennsylvania Law Review*, 151(3):1211–1254.
- Camerer, C. F. (2003). *Behavioral Game Theory*. Princeton University Press, Princeton, NJ.
- Charness, G., Gneezy, U., and Kuhn, M. A. (2013). Experimental methods: Extra-laboratory experiments—extending the reach of experimental economics. *Journal of Economic Behavior & Organization*, 91(0):93–100.
- Deaton, A. and Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine*, 210:2–21. Randomized Controlled Trials and Evidence-based Policy: A Multidisciplinary Dialogue.
- Dhami, S. (2016). *The Foundations of Behavioral Economic Analysis*. Oxford University Press, Oxford.
- Favereau, J. and Nagatsu, M. (2020). Holding back from theory: limits and methodological alternatives of randomized field experiments in development economics. *Journal of Economic Methodology*, 0(0):1–21.
- Fischbacher, U., Gächter, S., and Fehr, E. (2001). Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters*, 71:397–404.
- Gneezy, U., List, J., and Price, M. K. (2012). Toward an understanding of why people discriminate: Evidence from a series of natural field experiments. Working Paper 17855, National Bureau of Economic Research.
- Guala, F. (2005). *The Methodology of Experimental Economics*. Cambridge University Press, Cambridge, England.
- Guala, F. (2007). How to do things with experimental economics. In MacKenzie, D., Muniesa, F., and Siu, L., editors, *Do economists make markets? On the performativity of economics*, chapter 5, pages 128–162. Princeton University Press.
- Guala, F. (2009). Methodological issues in experimental design and interpretation. In Ross, D. and Kincaid, H., editors, *The Oxford handbook of philosophy of economics*. Oxford University Press.
- Guala, F. (2012). Experimentation in economics. In Mäki, U., editor, *Philosophy of Economics*, Handbook of the Philosophy of Science, pages 597–640. North-Holland, Amsterdam.
- Güth, W. and Kliemt, H. (2017). Experimental economics—a philosophical perspective. In *Oxford handbooks online*, pages 1–33. Oxford University Press.
- Harrison, G. W. (2019). The methodologies of behavioral econometrics. In Nagatsu, M. and Ruzzene, A., editors, *Contemporary philosophy and social science: An interdisciplinary dialogue*, chapter 4, pages 107–137. Bloomsbury Academic.
- Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4):1009–1055.
- Hertwig, R. and Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences*, 24:383–+.
- Jacquemet, N. and l’Haridon, O. (2018). *Experimental Economics*. Cambridge University Press.
- Jiménez-Buedo, M. (2011). Conceptual tools for assessing experiments: some well-entrenched confusions regarding the internal/external validity distinction. *Journal of Economic Methodology*, 18(3):271–282.
- Kahneman, D. (1988). Experimental economics: A psychological perspective. In Tietz, R., Albers, W., and Selten, R., editors, *Bounded Rational Behavior in Experimental Games and Markets*, pages 11–20, Berlin. Springer-Verlag.
- Kahneman, D. (2003). Biographical. Nobelprize.org.

- Kahneman, D., Knetsch, J. L., and Thaler, R. (1986a). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, 76(4):pp. 728–741.
- Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1986b). Fairness and the assumptions of economics. *The Journal of Business*, 59(4):pp. S285–S300.
- Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1990). Experimental tests of the endowment effect and the Coase theorem. *Journal of political Economy*, 98(6):1325–1348.
- Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic Perspectives*, 5(1):193–206.
- Lewin, S. B. (1996). Economics and psychology: Lessons for our own day from the early twentieth century. *Journal of Economic Literature*, 34(3):1293–1323.
- List, J. A. (2007). On the interpretation of giving in dictator games. *Journal of Political Economy*, 115(3):482–493.
- Mirowski, P. and Nik-Khah, E. (2007). Markets made flesh: performativity and a problem in science studies, augmented with consideration of the fcc auctions. In MacKenzie, D., Muniesa, F., and Siu, L., editors, *Do economists make markets? On the performativity of economics*, chapter 7, pages 190–224. Princeton University Press, Princeton, N.J.
- Moffatt, P. G. (2016). *Experimentics: Econometrics for experimental economics*. Palgrave, London.
- Muniesa, F. and Callon, M. (2007). Economic experiments and the construction of markets. In MacKenzie, D., Muniesa, F., and Siu, L., editors, *Do economists make markets? On the performativity of economics*, chapter 6, pages 163–189. Princeton University Press.
- Nagatsu, M. and Favereau, J. (2020). Two strands of field experiments in economics: A historical-methodological analysis. *Philosophy of the Social Sciences*, 50(1):45–77.
- Rabin, M. (1998). Psychology and economics. *Journal of Economic Literature*, 36(1):pp. 11–46.
- Rabin, M. (2002). A perspective on psychology and economics. *European Economic Review*, 46:657 – 685.
- Rapoport, A. and Seale, D. A. (2008). Coordination success in non-cooperative large group market entry games. In Plott, C. R. and Smith, V. L., editors, *Handbook of Experimental Economics Results*, volume 1, chapter 34, pages 273–295. Elsevier.
- Redpath, S. M., Keane, A., Andrn, H., Baynham-Herd, Z., Bunnefeld, N., Duthie, A. B., Frank, J., Garcia, C. A., Mnsson, J., Nilsson, L., Pollard, C. R. J., Rakotonarivo, O. S., Salk, C. F., and Travers, H. (2018). Games as tools to address conservation conflicts. *Trends in Ecology & Evolution*, 33(6):415–426.
- Reiss, J. (2018). Against external validity. *Synthese*.
- Roth, A. E. (1993). The early history of experimental economics. *Journal of the History of Economic Thought*, 15(2):184–209.
- Rubin, H., O’Connor, C., and Bruner, J. (2019). Experimental economics for philosophers. In Fischer, E. and Curtis, M., editors, *Methodological advances in experimental philosophy*, Advances in experimental philosophy, pages 175–206. Bloomsbury Publishing.
- Simon, H. A. (1986). Rationality in psychology and economics. *The Journal of Business*, 59(4):S209–S224.
- Smith, V. L. (1991). Rational choice: The contrast between economics and psychology. *Journal of Political Economy*, 99(4):877–897.
- Sugden, R. (2005). Experiments as exhibits and experiments as tests. *Journal of Economic Methodology*, 12(2):291–302.
- Van Egmond, S. and Zeiss, R. (2010). Modeling for policy science-based models as performative boundary objects for dutch policy making. *Science & Technology Studies*, 23(1):58–78.
- Varian, H. R. (2010). *Intermediate Microeconomics; a modern approach*. 8th edition.
- Worrapimphong, K., Gajaseeni, N., Le Page, C., and Bousquet, F. (2010). A companion modelling approach applied to fishery management. *Environmental Modelling & Software*, 25(11):1334–1344.