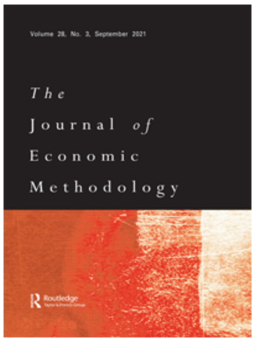

Journal of Economic Methodology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rjec20

Markets, market algorithms, and algorithmic bias

Philippe van Basshuysen

To cite this article: Philippe van Basshuysen (2022): Markets, market algorithms, and algorithmic bias, Journal of Economic Methodology, DOI: 10.1080/1350178X.2022.2100919

To link to this article: https://doi.org/10.1080/1350178X.2022.2100919

Published online: 01 Aug 2022.

JOURNAL OF ECONOMIC METHODOLOGY
https://doi.org/10.1080/1350178X.2022.2100919

Markets, market algorithms, and algorithmic bias

Philippe van Basshuysen [a,b]

[a]Leibniz University Hannover, Hanover, Germany; [b]London School of Economics and Political Science, London, UK

ABSTRACT
Where economists previously viewed the market as arising from a 'spontaneous order', antithetical to design, they now design markets to achieve specific purposes. This paper reconstructs how this change in what markets are and can do came about and considers some consequences. Two decisive developments in economic theory are identified: first, Hurwicz's view of institutions as mechanisms, which should be designed to align incentives with social goals; and second, the notion of market*places* – consisting of infrastructure and algorithms – which should be designed to exhibit stable properties. These developments have empowered economists to create marketplaces for specific purposes, by designing appropriate algorithms. I argue that this power to create marketplaces requires a shift in ethical reasoning, from *whether* markets should reach into certain spheres of life, to *how* market algorithms should be designed. I exemplify this shift, focusing on bias, and arguing that transparency should become a goal of market design.

ARTICLE HISTORY
Received 17 January 2022
Accepted 8 July 2022

KEYWORDS
Markets; algorithms; market design; bias; transparency

1. Introduction

Markets increasingly operate with formal rules, or *algorithms*, which can be designed to achieve a variety of objectives (Kominers et al., 2017; Milgrom, 2017; Roth, 2018). On the face of it, this notion may seem surprising. Many political economists have historically considered the market to be a *spontaneous order*: a state of social coordination, which has grown informally out of an interplay of individual actions, rather than being brought about by human design (e.g. Hayek, 1960, p. 160). This raises two questions: how did this change in what markets are and can do come about, and what are the consequences of this change?

To answer the first question, this paper offers a reconstruction of an episode in recent economic history, which emerged from the socialist calculation debate, proceeded to mechanism design and eventually evolved into the applied market design (Section 2). My reconstruction identifies two steps that were decisive in superseding the spontaneous-order view and establishing the new notion of what markets are and can do. The first step is Leonid Hurwicz' view of an institution as a mechanism, which translates individual actions into social outcomes, and that should be designed to align individual incentives with social goals (1973). While this view reconciled the market (and institutions more generally) with design, a further step was required for practical design efforts to succeed. This consisted in the notion of a market*place*, consisting of infrastructure and algorithms. Because marketplaces are parts of larger economic environments, which can interfere with their functioning, the algorithms governing a marketplace should be designed in a way that reduces possible

CONTACT Philippe van Basshuysen p.c.van-basshuysen@lse.ac.uk

'disturbances' from outside (Milgrom, 2021, p. 1384; see also Roth & Wilson, 2019). Using the redesign of the medical match as a case study, I shall argue that this step enabled economists to design market algorithms to implement a variety of social goals.

This 'power of market design' (Milgrom & Battilana, 2021), however, raises novel questions about how marketplaces ought, or ought not, to be designed, which did not arise under the spontaneous-order view. Thus, I shall argue that this development requires a shift in ethical reasoning about markets, from *whether* markets should reach into certain spheres of life, to *how* market algorithms ought to be designed (Section 3). In particular, I shall explore whether the design of market algorithms entails the risk that some market algorithms might be *biased* in problematic ways, unfairly advantaging certain groups or companies at the expense of others. Focusing again on the medical match, some ways are identified in which bias can be found and mitigated. Since the success of design may, however, depend not only on an algorithm being unbiased, but also on the trust of market participants in the fairness of the procedure, this highlights an additional problem, namely, whether a market algorithm is sufficiently transparent. Finally, I shall introduce another example – the colossal incentive auction of the radio spectrum – to show that opacity poses a severe obstacle to the scrutiny of market algorithms, suggesting that transparency should become an explicit goal of market design (Section 4).

2. What are markets, what can they do? From Hayek to Hurwicz to Roth

Friedrich Hayek distinguished three kinds of phenomena: natural phenomena that are independent of human action, artificial phenomena that are the result of human design, and 'a distinct middle category comprising all those unintended patterns and regularities which we find to exist in human society' (1967, p. 97). He called the third category, 'spontaneous order', and interpreted the market as exhibiting spontaneous order of a particular kind, in which prices efficiently coordinate private information and individual actions. He argued, further, that human design cannot replicate, much less improve upon, this spontaneous order (1945). While this argument was subject of fierce debate (which we will encounter presently), his notion of the market as a spontaneous order was shared by many of his contemporaries. More recently, however, the concept of the market has shifted from spontaneous to artificial: economists now recognize a variety of different marketplaces, which operate with algorithms that must be well-designed in order for a marketplace to be functional. Economists who design marketplaces accordingly interpret their work, in Alvin Roth's words, as 'engineering' (2002), bringing about desirable outcomes through appropriate design. This section reconstructs the change of how markets are viewed and what they can achieve as a two-step process – from the socialist calculation debate, via mechanism design, to applied market design.

2.1 From the socialist calculation debate to mechanism design

In the socialist calculation debate, economists quarreled over the relative merits of centrally planned versus market economies. On the one hand were market socialists like Oskar Lange, Abba Lerner, and Fred M. Taylor, who argued that a centrally planned economy could in principle replicate the efficient allocation of resources in a competitive market, and could improve on the workings of the market by correcting market failures (Lange & Taylor, 1938; Lerner, 1944). Notable opponents were Hayek and Ludwig von Mises, who fiercely disagreed with this thesis. According to Hayek, central planning could lead to efficient resource allocation only if the planner possessed at least as much information about the desires and resources of the individuals in the economy as the market mechanism generates spontaneously. He argued that this information is too dispersed for anyone to collect and oversee in its totality, and it is not in the interests of individuals to reveal their private information (1945). The market, on the other hand, spontaneously integrates this information through the prices that are generated by supply and demand. He concluded that it is impossible to rationally design an institutional arrangement that could simulate, let alone improve upon, the spontaneous order of the market.

Hayek's argument became widely influential (not only among economists in the Austrian tradition but beyond intradisciplinary borders), thus consolidating the concept of the market as a generic, spontaneous order in which the price system efficiently coordinates private information. However, the argument remained, in Roger Myerson's words, 'intuitive' (2008, p. 599; see also Caldwell, 2004, p. 217); the economic models available at the time of the debate accounted for economic systems only as mechanisms for the allocation of scarce resources, but not as mechanisms for communicating private information that is widely dispersed throughout the economy. These models thus ignored the informational properties of different economic systems and the ways in which different systems incentivize information sharing. In the absence of a mathematical treatment of information, a precise evaluation of Hayek's argument could not be conducted, and its validity remained inconclusive (Myerson 2008, 2009). Consequently, according to many observers, the calculation debate resulted in a draw (Caldwell, 2004, p. 338).

Hurwicz eventually broke this deadlock by introducing incentive constraints into economic models. In (1972), he modeled institutions as mechanisms, which determine how social decisions and the allocation of goods should depend on individuals' actions. Importantly, their actions include conveying private information, e.g. about their endowments, or their preferences. It is assumed that the agents are rational and game theory is used to predict institutional outcomes. In this framework, agents may act strategically, while adhering to the rules of the game. For instance, they may lie about their preferences if this is in their best interest. This leads to the concept of *incentive compatibility*: a mechanism is incentive-compatible if it implements some predefined social goal in equilibrium, that is, it gives everyone incentives to act according to the social plan. This analysis of incentives laid the foundations of mechanism design theory, which provides flexible tools for comparing the incentives and informational efficiency (i.e. how much agents need to know in order to achieve optimal results) that different mechanisms provide. For instance, Eric Maskin (2015) shows that Hayek's principal claims – that the market mechanism is the only informationally efficient and incentive-compatible mechanism – are true for a model in which there are large numbers of buyers and sellers and no externalities. However, if these assumptions are not met, there are generally mechanisms that improve upon the market.

Note that the market concept of early mechanism design is still a generic one, because *the market* can be modeled as a specific mechanism in which an interplay of individual actions produces prices and trades at those prices. This conception does not recognize a variety of different marketplaces. Hurwicz's theory did allow, however, for devising different economic systems as mechanisms and comparing their incentives. He expressed this idea by treating systems as variables whose value we seek to optimize, where the value could be the market or other institutional arrangements. For instance, he begins his 1973 Richard Ely Lecture, 'The Design of Mechanisms for Resource Allocation', by describing how mechanism design differs from more traditional economic analyses:

> Traditionally, economic analysis treats the economic system as one of the givens. The term 'design' in the title is meant to stress that the structure of the economic system is to be regarded as an unknown. An unknown in what problem? Typically, that of finding a system that would be, in a sense to be specified, superior to the existing one. The idea of searching for a better system is at least as ancient as Plato's Republic, but it is only recently that tools have become available for a systematic, analytical approach to such search procedures. This new approach refuses to accept the institutional status quo of a particular time and place as the only legitimate object of interest and yet recognizes constraints that disqualify naive utopias. (1973, p. 1)

Hurwicz thus fundamentally disagreed with Hayek's assertion that a superior system to the existing one cannot be designed, claiming that his novel theory could achieve precisely this. Furthermore, the market system here is not 'special' in the way it was for Hayek: the system to be designed could well be a market system. Hurwicz's market concept is thus compatible with the notion of designed markets.[1] Because of this disagreement, the relationship between Hayek's work and mechanism design is peculiar. On the one hand, mechanism design can be taken as emerging from Hayek's treatment of information and incentives. As we have seen above, Hayek seems to have anticipated crucial results that could later be proven within mechanism design theory, viz. that the market mechanism is the only informationally efficient and incentive-compatible mechanism. On the other hand, treating economic systems as variables in the problem of finding a superior system to the existing one must have appeared utterly unacceptable to Hayek, because it would undermine the spontaneous order of the market.

Though Hurwicz drew attention to the potential practical importance of his research program in allowing the design of alternatives to existing institutions, the program did not directly lead to real-world social reforms. An explanation may be found in Hurwicz' writings: 'The new mechanisms are somewhat like synthetic chemicals; even if not usable for practical purposes, they can be studied in pure form and so contribute to our understanding of the difficulties and potentialities of design' (1973, p. 27). Thus, while we can study the mechanisms to learn something about the design of institutions, they should not (yet) be implemented in the real world, perhaps in virtue of their 'pure form'. By this, he might have meant idealizations in game-theoretical models, such as stringent assumptions concerning agents' rationality and knowledge, which might not hold in the real world; or that agents might have additional choices that are omitted in a model, which might impair inferences from the model to the real world. Let us next consider how these problems were addressed as 'pure-form' mechanism design evolved into more small-scale, practical market design.

2.2 From mechanism to practical design

Makowski and Ostroy (1992) described the significance of Hurwicz's contribution thus: 'around 1970, the issue of incentives surfaced forcefully, as if a pair of blinders were removed' (p. 14). Following his breakthrough, any possible institutional arrangement could in principle be modeled as a mechanism, and the incentives that it provides to individuals could in principle be assessed. This possibility led to the rapid development of mechanism design and to a refinement of the theory. In this process, the theory branched into two broad research programs (Maskin, 2008): the first studies very general questions, such as how the set of feasible social outcomes can be characterized; while the second seeks to answer questions that arise in specific settings, such as how to design particular auctions, contracts, or matching programs.

The research on specific settings became an important factor in the emergence of practical market design. From the 1980s onwards, some of this research applied insights from mechanism design to real-world institutions, which would eventually arouse the interest of policy makers. For an illustration of this line of work, let's introduce an example, which will also serve as a basis for our discussion about bias in the next section: the National Resident Matching Program (NRMP), which assigns medical graduates to training positions ('residencies') in U.S. hospitals. This program had originated as a decentralized labor market at the beginning of the twentieth century, but as a response to market failures, a centralized matching procedure was introduced that took as input applicants' and hospitals' stated preferences over possible matches. This matching procedure worked to the satisfaction of applicants and hospitals for a while, but it was failing again when Roth (1984) studied it from the perspective of matching theory. David Gale and Lloyd S. Shapley had introduced some key notions of the theory in their 'stable marriage problem':

> A certain community consists of *n* men and *n* women. Each person ranks those of the opposite sex in accordance with his or her preferences for a marriage partner. We seek a satisfactory way of marrying off all members of the community … we call a set of marriages *unstable* (and here the suitability of the term is quite clear) if under it there are a man and a woman who are not married to each other but prefer each other to their actual mates. (Gale & Shapley, 1962, p. 388, emphasis in original)

Gale and Shapley formalized the marriage problem and defined a simple procedure in which men iteratively propose to their most preferred women, who then reject all but their most preferred proposal in each step. They showed that this 'deferred acceptance algorithm' produces stable matchings for every instance of the marriage problem, even in variations in which the numbers of men and women are unequal and one side of the market may be polygamous. How would this result relate to the medical match? The algorithm used in the medical match could be shown to be equivalent to a deferred acceptance algorithm where the hospitals ('men') propose to the applicants ('women') who accept or reject and the hospitals can be polygamous (viz., offer more than one position) (Roth, 1984). So this algorithm could (at least initially) be expected to generate stable outcomes. Roth noted, however, that the system started failing when increasing numbers of married couples entered the labor market, who may have complementary preferences (e.g. 'I prefer Boston to Chicago, but if my partner is assigned to the Great Lakes region my preferences switch'). His explanation of the novel failure was that such preferences can give rise to unstable matchings, and that this caused the unraveling of the medical match as increasing numbers of couples entered the market.

When the unraveling of the program was at a peak in the 1990s, the NRMP directors commissioned Roth with a redesign of the matching algorithm to restore an orderly procedure. The ensuing design process exhibits the ways in which practical design, or 'engineering' (Roth, 2002), goes beyond conventional mechanism design. In abstract matching models, unstable matchings are unsatisfactory because agents in blocking pairs have incentives to rematch outside the system. However, this result relies on stringent common-knowledge assumptions about the agents' preferences, which are not satisfied in the medical match, where neither applicants nor hospitals see the preference lists of others. So was a lack of stability really the feature that caused the failure of the program? An explanation for why unstable matchings could have led to market failure is that dissatisfied applicants might find out whether they were part of a blocking pair by contacting the hospitals that they preferred to their actual match. In this way, some applicants could identify more preferred hospitals that would be willing to hire them instead of some of their assigned applicants. Over time, this would lead to applicants and hospitals increasingly making deals on their own, instead of adhering to their assigned matches by the NRMP, which may have eventually caused failure of the program. Indeed, field and lab experiments were conducted, which confirmed that stability is the decisive factor for the functioning of matching programs in similar settings (see Roth, 2002).

As this case illustrates, an important step for the possibility of practical design was to think of institutions as market*places*, which are parts of larger economic environments (Roth, 2018). Marketplaces consist of infrastructure and *algorithms* – such as deferred acceptance algorithms in the case of matching programs – that are designed and implemented to achieve set goals (Kominers et al., 2017; Roth, 2018).[2] In order for the design to be successful, however, the designer must pay attention to the fact that marketplaces may interact with the bigger economic environments they are part of, in ways that game-theoretical models won't predict. In particular, people may have options outside the marketplace that a model depicts, for instance, in the medical match, doctors might make phone calls to check for other potential matches if they are unhappy with their assigned match. If a marketplace algorithm does not provide appropriate incentives, people might transact elsewhere, which might in turn cause failure of the marketplace. Thus, *a designed algorithm should not only be incentive-compatible concerning the strategies that are explicitly modeled (Hurwicz's lesson), but should also be stable concerning strategies that are not part of the model* (Milgrom, 2021; Roth & Wilson, 2019).

The case also shows that economic engineers have assumed the challenge of dealing with real people, rather than ideal economic agents (below, it will be argued that this brings into focus the importance of transparency). Idealized assumptions, such as common-knowledge assumptions in game-theoretical models, are thus viewed with suspicion, as is the assumption of the designer's full knowledge of players' strategies and information (Roth & Wilson, 2019). Applied market design thus cannot rely on the results of models alone; rather, experiments and computation are used to provide evidence that model results hold water in the messy, real world (Roth, 2002).[3]

Experiments and computation also paved the way for the redesign of the algorithm for the medical match. Although stable matchings cannot be guaranteed in a matching market with couples, computations showed that, because this marketplace exhibited certain features (e.g. that preference lists are relatively short), stable matchings exist with a high probability, even when the simple deferred acceptance algorithm in use would not find them. Roth and his colleagues thus designed a modified deferred acceptance algorithm in which applicants propose to hospitals, that seeks to find stable matchings by detecting blocking pairs and removing them at intermediate steps (Roth & Peranson, 1999). This algorithm was first implemented in 1998 and has since then created satisfactory outcomes. The redesign of the medical match and other cases of economic engineering – including the ambitious incentive auction of radio spectrum in the U.S. (Milgrom, 2017), which we will encounter below – have been celebrated as demonstrating the 'power and potential of market design to create and refine marketplaces' (Milgrom & Battilana, 2021). The shift from 'spontaneous' to 'designed' is thus completed.[4]

3. Markets, morals, and bias

Having understood how the change from the view of the market as a spontaneous order to the design of marketplaces came about, we are now in a position to appreciate the ethical implications. Many existing debates concerning markets and morals are reminiscent of the spontaneous-order view. These debates have tended to focus on *whether* markets should reach into certain spheres of life, and what the moral limits of markets are. Thus, some theorists in these debates argue that markets should be banned from certain spheres, troubled by the prospect that market norms could cover deeply human affairs such as love, or that human organs or blood could be for sale (e.g. Anderson, 1995; Sandel, 2012; Satz, 2010; Titmuss, 1970); while other theorists, perhaps influenced by the Hayekian notion of market superiority, would prefer to deregulate markets and let the spontaneous order work its wonders (e.g. Arrow, 1972; Narveson, 2001; Nozick, 1974). But these debates leave important questions unanswered from the perspective of the design view. The capacity of economic engineers to design marketplaces to achieve specific objectives, in particular, raises the question, which did not arise under the spontaneous-order view, of *how* these marketplaces ought, or ought not to be designed (cf. Brennan & Jaworski, 2016; Jaworski & Brennan, 2015).[5]

Shengwu Li has recently proposed a division of labor between market designers and moral philosophers (2017), which may be usefully applied to answer this question. Market designers, according to Li, should propose a variety of designs that speak to different ethical views, such as preference utilitarianism, or Kantian ethics, thus 'maintain[ing] an informed neutrality between reasonable ethical positions' (2017, p. 717). They should leave it up to philosophers and policy makers to debate substantive moral questions, about what values and objectives should guide design (in general, or in a given case), and how trade-offs should be decided when not all objectives can be simultaneously achieved. The envisioned division of labor presupposes that moral philosophers be acquainted with the broad methodology of market design and with some of its results (van Basshuysen, 2019); when assessing a matching market, for instance, they should know how fairness is defined in this context, and whether it is compatible with efficiency, as it would be unreasonable to demand mutually incompatible objectives.

This general method, if followed, could ensure that particular designs are evaluated and chosen while accounting for moral values. There may, moreover, be more general ethical questions that arise over a variety of designs.[6] In the remainder of this paper, I shall consider some novel questions that may arise in various design contexts, focusing on 'market algorithms', i.e. the algorithms that are designed and implemented to facilitate marketplaces. Might some market algorithms be *biased* in problematic ways, unfairly disadvantaging some groups of market participants in favor of another group? And if so, how could such a bias be eliminated and a marketplace be made more equitable? While algorithmic bias has recently received increasing attention (e.g. Hedden, 2021), market algorithms have been largely neglected in this debate thus far.[7] So let's first consider whether there are reasons to believe that a problematic kind of bias might not arise in the context of market algorithms, i.e. if there are reasons that would justify this neglect. In the next step, we will consider the questions of what bias might amount to in this context.

Note, first, that not every kind of bias raises fairness issues; for instance, a person would be regarded as a bad parent if they were not biased towards their children. One reason to believe that, even when a market algorithm is biased, this might not be problematic, is that markets typically concern the trading of private property, where fairness is not usually regarded as a relevant issue. However, this is not in general a compelling argument when it comes to market design. As we have seen, marketplaces may be designed for employment, or for public resources such as spectrum, and thus they often concern 'situations that were not previously thought of as transactional' (Li, 2017, p. 718), and where a systematic bias against certain groups would be regarded as problematic. A second reason to think that biased market algorithms might not be problematic is the view that individuals interacting through markets freely agree to the terms of the deal and will (at least when the market is competitive) have alternative options available to them. Thus, if outcomes on a marketplace are biased against them, they could simply go on to transact elsewhere. However, the methodology of market design, as described in Section 2, should make us wary of this argument. As we have seen, market designers take pains to create algorithms with stable properties, thus making it preferable for participants to stick to the marketplace and preventing them from taking outside options. It follows that, even when participants are systematically disadvantaged by a market algorithm, they may be even worse off when leaving this marketplace.[8] Because such an algorithm will remove true alternatives, it should be regarded as problematic if it is biased.

Having found the reasons to believe that we need not worry about biased market algorithms unconvincing, let us examine how bias may arise and how it might be overcome, using matching algorithms as an example. For instance, in simple matching markets,[9] a deferred acceptance algorithm in which employers propose to applicants produces stable matchings that are weakly preferred by all employers to all other stable matchings, and weakly dispreferred by all applicants to all other stable matchings, and vice versa for an applicant-proposing deferred acceptance algorithm (Gale & Shapley, 1962). Thus, the use of deferred acceptance algorithms tends to favor the proposing side of the market over the side receiving the proposals and can thus be regarded as biased.

How might this turn out in practice? Friedman and Nissenbaum, focusing on the medical match before its redesign, argue that the algorithm that was then in use unfairly favored hospitals over applicants, and among the applicants, unfairly disadvantaged married couples (1996). They suggest that the bias against applicants was due to the fact that the algorithm was hospital-proposing; and that the bias against couples was due to the unstable matchings that the algorithm would generate (which would produce blocking pairs especially among couples). As it will turn out, the redesign of the algorithm successfully removed these biases, and thus provides a good case study for mitigation strategies. We saw earlier that, even though there is no guarantee that stable matchings exist when couples are present, the medical match has certain features that make it likely that stable matchings exist, and an algorithm was designed that would seek to find them by removing blocking pairs. This would thus eliminate one kind of bias identified above, that is, the bias against couples. But what about the other bias – how well off would applicants be compared to hospitals? Here, computational experiments suggested another, important fact, namely that the set of stable matchings, while non-empty, would be *small* (Roth & Peranson, 1999). This is important because, as the set of stable matchings becomes small, an algorithm producing stable matchings will be free of bias, because there will be very few applicants and hospitals that are matched differently under different stable matchings. Thus, the new algorithm, which was designed to find stable matchings in this market, can be regarded as practically unbiased.[10]

The fact that debiasing was possible in the medical match, of course, does not mean things always work out so well: in many settings that differ structurally (for instance in simple matching markets without couples), the set of stable matchings may be large, thus permitting a lot of 'room' for bias, since many agents will be matched differently under different stable matchings. In these cases, it will be impossible to design an algorithm that does not favor one side of the market over another, or some group of market participants over others, which might be regarded as unfair when there is no reason to favor particular groups over others.[11] In cases where bias is inevitable, however, it may still be possible to achieve a kind of procedural fairness: these are probabilistic algorithms which, while not mitigating the bias in the end state, are *ex ante* fair, that is, market participants will have the same chance of coming out on the advantaged side of the market (Klaus & Klijn, 2006).

These debiasing strategies, however, face an additional challenge: their success is contingent on the trust of market participants in the fairness of a procedure. It is crucial here that, as we saw in the previous section, designed marketplaces are populated by real people, rather than ideal economic agents. This matters for the design of market algorithms; for even when a designed algorithm is unbiased, it does not follow that market participants know, or believe that it is. At the same time, their belief in the fairness of the procedure may be crucial for the acceptance and social support of a marketplace, a lack of which may in turn lead to market failure (see Roth, 2007). So how can we ensure people trust that an algorithm is fair? An obvious answer is that they will trust in its fairness when they understand how the algorithm operates, or what kinds of outcomes it tends to generate, which, at least in some cases, can be conveyed to market participants. The redesign of the medical match can, again, be seen as a role model in this regard. Here, the designers took pains to explain the implications of the fact that the set of stable matchings tends to be small, in particular convincing hospitals that the proposed algorithm wasn't biased against them (even though this algorithm is essentially applicant-proposing) (Roth, 2002). Consequently, the proposed algorithm did not face opposition and has been regarded as well-behaved since its implementation. But other algorithms may resist scrutiny by market participants, or even by impartial observers. We now turn to this problem.

4. Opacity

When a market algorithm is transparent, both the nature of the possible bias and its source can be clearly located and anticipated – this is the case, for instance, for deferred acceptance algorithms in matching markets, which can easily be written out on a piece of paper and its properties understood. Other algorithms, however, are more epistemically opaque, making it difficult or impossible for stakeholders or outside observers to understand relevant features of these algorithms and what kinds of outcomes they tend to generate. The problem is that, when such an algorithm is biased, it may, due to its opacity, be difficult to detect and mitigate this bias.

Opacity may be attributed to a range of factors, such as difficulties in translating policy goals into formal criteria, the complexity of an allocation problem, lack of relevant data, informational issues, or strategic behavior of market participants, as we shall see below.[12] It may be futile to ask for a precise threshold for opacity, not only because opacity can come in these different varieties, but also because, whether an algorithm is regarded as opaque will depend on *whom* a marketplace is designed for; for instance, while some auction might appear highly complex to non-expert bidders, the same auction might be regarded as transparent if the bidders are firms that, perhaps, draw on the expertise of game theorists.[13] There are, nevertheless, unambiguous cases and it will be useful here to introduce an example that may be regarded as particularly opaque, and which has (partly for this reason) attracted considerable attention: the *incentive auction*, conducted by the Federal Communications Commission (FCC) in 2016/17, as a response to an altered demand for spectrum usage from television to mobile phones. Milgrom et al. (2012) had proposed an incentive auction as an efficient way to reallocate spectrum from TV broadcasters to mobile broadband companies. The auction has a complex design which, roughly, consists of a reverse auction determining the price at which broadcasters would give up their spectrum, a 'repacking' process in which broadcasters retaining their spectrum would be assigned new channels, and a forward auction determining the price that mobile companies were willing to pay for spectrum usage. The repacking process especially was subject to highly complex constraints whose solution was computationally hard (Milgrom, 2017; Milgrom & Segal, 2020).

While the incentive auction was generally hailed as a success, some concerns were also raised, which can be linked to its opacity. Doraszelski et al. simulated the auction and found that the repacking process could enable market manipulations, making it a profitable strategy for some bidders in the reverse auctions to strategically withhold some TV stations in order to increase the prices of some of the other stations they were also holding (2019). They pointed out that these findings are consistent with the behavior of various private equity firms, which had acquired TV stations prior to the auctions in a way that would make such strategic supply reduction possible. They also estimated that the scale of these manipulations was large, transferring wealth from government to private equity firms, and potentially decreasing the efficiency of the auction and reducing the amount of repurposed spectrum.

Based on these results, should the incentive auction be regarded as exhibiting a problematic bias towards private equity firms? Some have interpreted the auction in this vein, arguing that it had been designed in such a way that raised too much money for private equity firms and too little for the government, thus transferring wealth from taxpayers to those firms (see Weyl, 2020). The claim that the auction was biased is hard to evaluate however as, due to the opacity of this case, the real scale at which manipulations occurred in the auction remains unclear. For not only was the allocation problem highly complex; in addition, the opacity was amplified here by the fact that the FCC had not released data on actual bidding behavior by the time Doraszelski et al. were investigating the possibility of manipulations. This meant, they had to resort to estimated reservation values of TV stations from secondary data sources to project the size of the manipulations.[14] How prevalent manipulations really were thus remains unclear to this date, as does the counterfactual question of whether viable alternative designs could have mitigated the possibilities for strategic supply reduction.

Thus, the hard problem in scrutinizing market algorithms is opacity, because this makes it difficult to exclude the possibility that there is a bias, and even more difficult to precisely locate the possible bias and to detect debiasing strategies. This suggests that transparency should be regarded as a design goal itself. It may not always be possible to achieve complete transparency, and there might be tradeoffs, for instance, between transparency and efficiency, or between transparency and market size (cf. Asquith et al., 2013). Nevertheless, a failure to achieve transparency is not only ethically problematic; it may also have detrimental practical effects because, once it is understood that the design of market algorithms is a human activity, in which vested interests may be at stake, questions might arise about the designers' own interests in a design. Indeed, some of the economists involved in the design of the incentive auction have been publicly accused of having a personal stake in the matter. The ensuing debates have been fierce,[15] as may be expected when the politics of market design stands as a proxy for possible issues with a market algorithm. To ensure more constructive debates over bias, it seems crucial to focus on transparency, and to regard this as an explicit goal of market design.

5. Conclusion

Political economists previously considered the market to be a kind of spontaneous order – a notion that is perhaps most clearly exhibited in the work of Hayek. Nowadays, however, they design algorithms inducing marketplaces to fulfill a variety of purposes. I identified two steps in economic theory – Hurwicz's view of institutions as mechanisms, and the notion of marketplaces that interact with larger economic environments – which jointly brought about this new notion of what markets are and can do. Since this development in economic theory has entailed an increased power of economists to shape the social world in a way that they, or their customers, deem desirable, this brings into focus the possibility that some designed market algorithms might (whether intentionally or not) generate outcomes that put some individuals or groups at an unfair disadvantage. Using the example of matching markets, I discussed some ways in which bias might be identified and mitigated. Because the success of design may, however, depend not only on an algorithm being unbiased, but also on people's trust in the fairness of the procedure, transparency becomes central. I therefore suggested this might be investigated as an explicit goal of market design. More generally, it is crucial that philosophers do not lose sight of these developments in economic theory, putting issues of justice and fairness on the market.

Notes

1. It is possible that for Hurwicz the market could sometimes also be a spontaneous order that developed in the absence of design, but the concept of spontaneous order does not appear to play an important role in his writings.
2. Roth (2018) adds that customs are also part of marketplaces, as local customs and norms may affect outcomes (see pp. 1610 and 1615). For our purposes, customs can be ignored.
3. See van Basshuysen (2021) for an attempt at providing a general methodology of market design.
4. It might be objected here that what is designed are not markets but mere market*places*, and therefore market design is presumably consistent with the view of the market as a spontaneous order. However, since marketplaces are parts of, embedded within, larger markets, the latter are, at least partly, subject of design too. I thank an anonymous reviewer for pressing me on this point.
5. Brennan and Jaworski argue that especially anti-commodification theorists have tended to overlook that their arguments concern contingent features of particular markets, rather than the general features shared by all kinds of markets. This seems to motivate a similar claim: rather than debating the reach of 'the market', moral philosophers should focus on the ways in which particular markets are designed.
6. An example of an evaluation of a specific design is van Basshuysen (2020) who argues that removing barriers to kidney exchange programs is a moral imperative. Hitzig (2020) seeks to establish general claims, arguing that design economics should allow for democratic participation.
7. An exception is a paper by Friedman and Nissenbaum who discuss possible bias in the context of the medical match prior to its redesign (1996). We shall encounter their arguments presently.
8. This problem may be particularly severe in two-sided marketplaces, such as matching markets and some auctions, where, in Roth's dictum, 'you can't just choose what you want, even if you can afford it: you also must be chosen' (Roth, 2018, p. 1612). If an algorithm is biased in favor of one side in such a marketplace, members of this side will stick to the marketplace that tends to favor them, and will thereby eliminate outside options for members of the disadvantaged side.
9. These are matching markets, such as in Gale and Shapley's marriage problem, in which there are no couples.
10. The algorithm also practically makes it a dominant strategy for applicants and programs to state their true preferences, making unlikely a different possible source of bias, namely bias due to strategic behavior (see below).
11. Technically speaking, the impossibility of finding unbiased stable matchings when the set of stable matchings is large is due to the fact that this set is a distributive lattice (Knuth, 1996).
12. Cf. Kominers and Teytelboym (2020) who discuss a subset of these factors as kinds of *complexity* that designers must overcome to achieve their goals. It is natural to think that the complexity of a design task can add to the epistemic opacity of the designed algorithm.
13. That opacity is relative to a cognitive agent is a general feature, see Humphreys (2009).
14. These data have in the meantime been released, but it appears that their implications have not yet been studied in detail.
15. Feltri (2020) suggested there were possible conflicts of interest arising from the academic and business relations between various market designers, including Paul Milgrom. Milgrom (2020) responded to both Weyl (2020) and Feltri (2020), claiming that the allegation that the auction favored private equity firms was unfounded, and rejecting the allegations that there were problematic relations among the designers.

Acknowledgements

The author thanks Luc Bovens, Elinor Clark, Enno Fischer, Jurgis Karpus, Donal Khosrowi, Sebastian Krug, Chrys Mantzavinos, Luis Enrique Segoviano Contreras, Lucie White, and Jannik Zeiser for helpful comments and discussions at various stages of this research. The author also thanks the audiences at the following events: the Soul of Economics conference, University of Zurich, 2019; the International Research Network Kick-Off Meeting 'Justice and Interest', Aix-Marseille School of Economics, 2019; the Philosophy of Economics Research Seminar, UNAM, Mexico City, 2020; the Philosophy of Social Science Roundtable, Emory University, Atlanta, 2020; and the 15th Biennial Meeting of the International

Network of Economic Method (INEM 2021) in Tempe, Arizona. Financial support from the VolkswagenStiftung within the project 'Bias and Discrimination in Big Data and Algorithmic Processing: Philosophical Assessments, Legal Dimensions, and Technical Solutions' is gratefully acknowledged.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Volkswagen Foundation.

Notes on contributor

Philippe van Basshuysen is a postdoctoral researcher at the Leibniz University Hannover and research associate at the Centre for the Philosophy of Natural and Social Science at the London School of Economics. His research is in the philosophy of economics and public policy, with a focus on public health.

ORCID

Philippe van Basshuysen http://orcid.org/0000-0003-1947-9309

References

Anderson, E. (1995). *Value in Ethics and economics*. Harvard University Press.

Arrow, K. J. (1972). Gifts and exchanges. *Philosophy & Public Affairs*, *1*(4), 343–362. https://doi.org/10.1515/9781400853564.139

Asquith, P., Covert, T., & Pathak, P. A. (2013). The effects of mandatory transparency in financial market design: evidence from the corporate bond market. *NBER Working Paper* 19417.

Brennan, J., & Jaworski, P. (2016). *Markets without limits: Moral virtues and commercial interests*. Routledge.

Caldwell, B. (2004). *Hayek's challenge: An intellectual biography of F.A. Hayek*. The University of Chicago Press.

Doraszelski, U., Seim, K., Sinkinson, M., & Wang, P. (2019). Ownership concentration and strategic supply reduction. *National Bureau of Economic Research Working Paper* 23034.

Feltri, S. (2020). It is such a small world: the market-design academic community evolved in a business network. *ProMarket*, 28 May. Retrieved September 14, 2021. https://promarket.org/2020/05/28/it-is-such-a-small-world-the-market-design-academic-community-evolved-in-a-business-network/.

Friedman, B., & Nissenbaum, H. (1996). Bias in computer systems. *ACM Transactions on Information Systems*, *14*(3), 330–347. https://doi.org/10.1145/230538.230561

Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *American Mathematical Monthly*, *69*(1), 9–15. https://doi.org/10.1080/00029890.1962.11989827

Hayek, F. A. (1945). The use of knowledge in society. *The American Economic Review*, *35*(4), 519–530.

Hayek, F. A. (1960). *The constitution of liberty*. University of Chicago Press.

Hayek, F. A. (1967). The results of human action but not of human design. In F. A. Hayek (Ed.), *Studies in philosophy, politics and economics* (pp. 96–105). University of Chicago Press.

Hedden, B. (2021). On statistical criteria of algorithmic fairness. *Philosophy & Public Affairs*, *49*(2), 209–231. https://doi.org/10.1111/papa.12189

Hitzig, Z. (2020). The normative gap: Mechanism design and ideal theories of justice. *Economics & Philosophy*, *36*(3), 407–434. https://doi.org/10.1017/S0266267119000270

Humphreys, P. (2009). The philosophical novelty of computer simulation models. *Synthese*, *169*(3), 615–626. https://doi.org/10.1007/s11229-008-9435-2

Hurwicz, L. (1972). On informationally Decentralized systems. In C. B. McGuire, & R. Radner (Eds.), *Decision and organization: A volume in honor of Jacob Marschak* (pp. 297–336). University of Minnesota Press.

Hurwicz, L. (1973). The design of mechanisms for resource allocation. *The American Economic Review*, *63*(2), 1–30.

Jaworski, P. M., & Brennan, J. (2015). Market architecture: It's the How. Not the What. *Georgetown Journal of Law & Public Policy*, *13*, 231–250.

Klaus, B., & Klijn, F. (2006). Procedurally fair and stable matching. *Economic Theory*, *27*(2), 431–447. https://doi.org/10.1007/s00199-004-0602-5

Knuth, D. E. (1996). Stable marriage and its relation to other combinatorial problems. *American Mathematical Society: CRM Proceedings and Lecture Notes, Volume 10*.
Kominers, S. D., & Teytelboym, A. (2020). The parable of the auctioneer: Complexity in Paul R. Milgrom's *discovering prices*. *Journal of Economic Literature*, *58*(4), 1180–1196. https://doi.org/10.1257/jel.20191504
Kominers, S. D., Teytelboym, A., & Crawford, V. P. (2017). An invitation to market design. *Oxford Review of Economic Policy*, *33*(4), 541–571. https://doi.org/10.1093/oxrep/grx063
Lange, O., & Taylor, F. M. (1938). In B. E. Lippincott (Ed.), *On the Economic theory of socialism*. University of Minnesota Press.
Lerner, A. (1944). *The economics of control: Principles of welfare economics*. Macmillan.
Li, S. (2017). Ethics and market design. *Oxford Review of Economic Policy*, *33*(4), 705–720. https://doi.org/10.1093/oxrep/grx047
Makowski, L., & Ostroy, J. M. (1992). General equilibrium and market socialism: Clarifying the logic of competitive markets. In P. Bardhan, & J. Roemer (Eds.), *Market socialism: the current debate*. Oxford University Press.
Maskin, E. S. (2008). Mechanism design: How to implement social goals. *The American Economic Review*, *98*(3), 567–576. https://doi.org/10.1257/aer.98.3.567
Maskin, E. S. (2015). Friedrich von Hayek and mechanism design. *Review of Austrian Economics*, *28*(3), 247–252. https://doi.org/10.1007/s11138-015-0310-3
Milgrom, P. (2017). *Discovering prices: Auction design in markets with complex constraints*. Columbia University Press.
Milgrom, P. (2020). The market design community and the broadcast incentive auction: fact-checking Glen Weyl's and Stefano Feltri's false claims. *Auctionomics*, 03 June. Retrieved on September 14, 2021. http://www.auctionomics.com/wp-content/uploads/2020/06/Fact-checking-Weyl-and-Feltri.2.pdf.
Milgrom, P. (2021). Auction research evolving: Theorems and market designs. *American Economic Review*, *111*(5), 1383–1405. https://doi.org/10.1257/aer.111.5.1383
Milgrom, P., Ausubel, L., Levin, J., & Segal, I. (2012). Incentive auction rules option and discussion. *Appendix C to the FCC's Notice of Proposed Rulemaking*, GN Docket no. 12-268, October 2.
Milgrom, P., & Battilana, S. C. (2021). The power of market design. *Project Syndicate*, Jul 15, Retrieved August 20, 2021. https://prosyn.org/k7b7CNN.
Milgrom, P., & Segal, I. (2020). Clock auctions and radio spectrum reallocation. *Journal of Political Economy*, *128*(1), 1–31. https://doi.org/10.1086/704074
Myerson, R. B. (2008). Perspectives on mechanism design in economic theory. *American Economic Review*, *98*(3), 586–603. https://doi.org/10.1257/aer.98.3.586
Myerson, R. B. (2009). Fundamental theory of institutions: A lecture in honor of Leo Hurwicz. *Review of Economic Design*, *13*(1-2), 59–75. https://doi.org/10.1007/s10058-008-0071-6
Narveson, J. (2001). *The libertarian idea*. Broadview Press.
Nozick, R. (1974). *Anarchy, State, and Utopia*. Basic Books.
Roth, A. E. (1984). The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy*, *92*(6), 991–1016. https://doi.org/10.1086/261272
Roth, A. E. (2002). The economist as engineer: Game Theory, experimentation, and computation as tools for design economics. *Econometrica*, *70*(4), 1341–1378. https://doi.org/10.1111/1468-0262.00335
Roth, A. E. (2007). Repugnance as a constraint on markets. *Journal of Economic Perspectives*, *21*(3), 37–58. https://doi.org/10.1257/jep.21.3.37
Roth, A. E. (2018). Marketplaces, markets, and market design. *American Economic Review*, *108*(7), 1609–1658. https://doi.org/10.1257/aer.108.7.1609
Roth, A. E., & Peranson, E. (1999). The redesign of the matching market for American Physicians: Some engineering aspects of economic design. *American Economic Review*, *89*(4), 748–780. https://doi.org/10.1257/aer.89.4.748
Roth, A. E., & Wilson, R. B. (2019). How market design emerged from Game Theory: A mutual interview. *Journal of Economic Perspectives*, *33*(3), 118–143. https://doi.org/10.1257/jep.33.3.118
Sandel, M. (2012). *What money can't buy: The moral limits of markets*. Farrar, Straus and Giroux.
Satz, D. (2010). *Why some things should not be for sale: The moral limits Of markets*. Oxford University Press.
Titmuss, R. M. (1970). *The gift relationship: From human blood to social policy*. Allen & Unwin.
van Basshuysen, P. (2019). *Philosophical and ethical aspects of economic design* [Unpublished doctoral dissertation]. London School of Economics and Political Science.
van Basshuysen, P. (2020). Kidney exchange and the ethics of giving. *Journal of Ethics and Social Philosophy*, *18*(1), 85–110. https://doi.org/10.26556/jesp.v18i1.895
van Basshuysen, P. (2021). How to build an institution. *Philosophy of the Social Sciences*, *51*(2), 215–238. https://doi.org/10.1177/0048393120971545
Weyl, G. (2020). How market design economists helped engineer a mass privatization of public resources. *ProMarket*, 28 May. Retrieved August 14, 2021. https://promarket.org/2020/05/28/how-market-design-economists-engineered-economists-helped-design-a-mass-privatization-of-public-resources/.