Article

How to Build an Institution

Philosophy of the Social Sciences
1–24

DOI: 10.1177/0048393120971545
journals.sagepub.com/home/pos

Philippe van Basshuysen[1]

Abstract

How should institutions be designed that "work" in bringing about desirable social outcomes? I study a case of successful institutional design—the redesign of the National Resident Matching Program—and argue that economists assume three roles when designing an institution, each of which complements the other two: first, the *designer* combines positive and normative modeling to formalize policy goals and to design possible mechanisms for bringing them about. Second, the *engineer* refines the design by conducting experiments and computational analyses. Third, the *plumber* implements the design in the real world and mends it as needed.

Keywords

economic design, engineering, plumbing, normative modeling, algorithmic bias

1. Introduction

Economists increasingly aspire to create or change institutions in order to bring about desirable social outcomes. While this field of economic design is growing, philosophers of social science have to date focused on a single case, namely the design of the early spectrum auctions conducted by the Federal Communications Commission (FCC) (Alexandrova 2008; Guala 2005). But by focusing on one case only, they have fallen short of providing a general

Received 11 October 2020

[1]Leibniz University Hannover, Hannover, Germany

Corresponding Author:
Philippe van Basshuysen, Leibniz University Hannover, Im Moore 21, 30167 Hannover, Germany.
Email: philippe.van.basshuysen@philos.uni-hannover.de

account of the practice of economic design. I aim, here, to provide a more general account, by introducing a case study that hasn't previously been considered in this literature: the *National Resident Matching Program* (NRMP), which places medical graduates in the US into training positions in hospitals. In the 1990s, economists guided the redesign of this labor market, which was commissioned as a response to severe market failures. The result is considered to be one of the foremost success stories of economic design, and a key motivation behind the Nobel Memorial Prize awarded to Lloyd Shapley and Alvin Roth in 2012.

Since the matching program and the spectrum auctions constitute the flagship cases of successful institutional design,[1] a general methodological account should be consistent with both cases. I shall thus put forward a methodology of economic design and of how knowledge is generated in the design process, which is consistent with the two design processes, and which combines insights from both. The account depicts economic design as a three-stage process in which the economist (be it one and the same or different ones dividing labor) assumes three roles.

The *economist-as-designer* (Hurwicz 1973) generates models in which properties corresponding to policy goals are defined, and mechanisms designed, that is, algorithms determining institutional outcomes for possible combinations of individual actions, which might bring about the defined goals if instituted in the real world. As will become clear, designers engage in a specific combination of positive and normative modeling, intervening in the models in order to learn how positive and normative constraints interact, and what the limits are to what could possibly be implemented.

The *economist-as-engineer* (Roth 2002) conducts experiments to examine whether the model results hold water, as well as computational analyses to quantitatively investigate the trade-offs and limits that the designer previously identified. Drawing on the results of these investigations, the engineer then makes proposals to refine the envisioned algorithm.

The *economist-as-plumber* (Duflo 2017) implements the designed algorithm in the real world, evaluates its functioning, and mends it if problems arise. In doing so, the plumber attends not only to the algorithm itself, but also to cultural and political issues surrounding the design, such as whether market participants trust that the designed algorithm is unbiased.

I shall argue that successful institutional design requires each of the three roles because they complement each other in important ways. Conversely, the

[1]The most recent Nobel Memorial Prize was awarded jointly to Paul R. Milgrom and Robert B. Wilson for their contributions to auction theory, including the application of the theory to the design of the FCC auctions (Committee for the Prize in Economic Sciences in Memory of Alfred Nobel 2020).

omission of any of these roles in the design process is likely to yield unreliable institutions that contain unintended biases.

The paper is structured as follows. In Section 2, I present some of the conclusions that philosophers of science have drawn from spectrum auction design concerning how models and experiments are used in economic design. Section 3 describes the redesign of the matching market for medical graduates in depth. In Section 4, I provide an account of economic design that is consistent both with the case of the spectrum auctions and the matching market. Section 5 concludes by proposing a normative reading of this account.

2. Philosophers of Science on Spectrum Auctions

In the most general terms, the aim of economic design is to bring about desirable social outcomes through the formulation of suitable institutional rules and infrastructure. The kinds of outcomes that are pursued may differ from case to case and may be contingent on political, economic and ethical considerations.[2] Although important, these are not the focus of this paper, and it will thus be assumed that exogenous policy goals completely determine the properties of desired outcomes, thus subsuming all the relevant normative considerations. Typically, designers seek to exploit agents' conscious, strategic pursuit of their individual goals in order to implement these properties, which explains why models from game theory, or rational choice models more generally, play an important role in design. This feature distinguishes institutional design from other types of policy designs such as nudges, which typically exploit agents' unconscious biases.

The case study of economic design discussed in the philosophical literature to date is the design of auctions for allocating spectrum licenses to telecommunication service providers in the US. Spectrum refers to a range of electromagnetic frequencies, which are used to transmit video, sound and data. In the 1990s, the FCC, an independent agency of the US government responsible for the allocation of the licenses, decided to replace an inefficient lottery system with auctions. The auctions were supposed to achieve specific policy goals, in particular efficiency, that is, to allocate the licenses to those providers that value them most. What kind of auctions would best promote these goals was subject to much controversy among the stakeholders, and the FCC as well as potential bidders consulted economists about crucial design decisions.

[2]See Li (2017) on the relationship between ethics and market design.

Since the first auctions were conducted in 1994, their design has widely been regarded as an efficient means of allocating licenses, and has raised billions of dollars in revenue for American taxpayers.[3] Moreover, it was supposedly economic theorists, in particular game theorists, who designed these auctions, so they were presented in media and, not surprisingly, by the theorists themselves, as an exemplar of the transformative force of game theory. For example, R. Preston McAfee and John McMillan, two theorist-consultants,[4] wrote: "*Fortune* said it was the 'most dramatic example of game theory's new power. . .It was a triumph, not only for the FCC and the taxpayers, but also for game theory (and game theorists)'" (McAfee and McMillan 1996, p. 159; in the quote, they refer to *Fortune magazine*, February 6, 1995, p. 3).

Philosophers of science have challenged this received view. Francesco Guala has convincingly argued that the successful design should not be credited to game theory alone, but rather emphasizes the role of laboratory experiments (see 2001, 2005, 2006, 2007). As he notes, no theorem from auction theory—a subfield of game theory—was directly applicable to the design of the auctions. The main problem was that the values that bidders attach to licenses often depend on whether they also get complementary licenses. For instance, these could be licenses in a neighboring state for a bidder who wishes to extend coverage. But different bidders may prefer different bundles of licenses, and thus it was not possible for the FCC to simply auction off all the complementary licenses as packages. Instead, the bundles were to be determined through the bidding process. However, there were no analytical solutions to what kinds of auction rules would achieve efficient allocations of goods that include complementarities. In particular, it was unclear whether bidders should be allowed to bid for licenses only individually or whether *package bidding* should also be allowed, in which bidders can submit single bids on packages of licenses. Both formats can give rise to problems: in individual auctions, bidders might win only a part of the package they would have liked to hold, which might reduce efficiency, while in package bidding auctions it might be intractable for the auctioneer to select winning bids (i.e., those that would maximize revenue) because of the large number of possible combinations of packages that bidders may bid on, which might also reduce efficiency. However, theory alone did not allow for the quantification and comparison of these problems, and thus fell short of deciding between auction formats.

[3]See Nik-Khah (2008) for a different interpretation.

[4]By the time, both were professors of economics, McAfee at the University of Texas, Austin, and McMillan at the University of California, San Diego. McAfee was involved in the design of the auctions working for the wireless telephone service provider AirTouch Communications, and McMillan for the FCC.

Complementarities were but one complication that analytical models alone could not resolve. Another was the "winner's curse": a phenomenon to be prevented, in which the bidder who most overestimates the value of a good wins the auction but thereby makes a loss. There was some evidence, both theoretical and experimental, that open instead of sealed-bid auctions could help reduce instances of the winner's curse (roughly, this is because bidders can learn about the true valuation by observing other bidders). However, whether this would turn out to be true in the presence of complementarities, and further complications of the market, was not clear. The problem was to find out whether and how these features would interact—and again, models from auction theory were silent on the matter.

Experimentalists were required to cope with these complications. Most prominently, Caltech economist Charles Plott was involved in the auctions, consulting for the telecom provider Pacific Bell. Together with his team, Plott created *experimental testbeds*; controlled laboratory environments of auctions in which some features, such as complementarities, can be controlled for. Importantly, testbeds test material environments holistically, allowing for the observation of possible interactions between different causal mechanisms. This distinguishes testbeds from models, which typically isolate particular causal mechanisms (Cartwright 2009; Mäki 2011). The experimentalists conducted numerous testbeds, which eventually favored simultaneous multiple-round auctions over alternative formats. These consist of rounds of open ascending bid auctions in which all the licenses are auctioned simultaneously, and where after each round the results are made public, so that bidders can gain information about their chances of assembling their preferred combinations of spectrum and accordingly adjust their bidding behavior. At the beginning, these auctions did not provide for package bidding, but, building on theoretical work by Lawrence Ausubel and Paul Milgrom (2002), since the mid-2000s, they have increasingly been combined with package bidding.

Anna Alexandrova and Robert Northcott have contrasted the ways in which theoretical economists (such as McMillan) and experimental economists (Plott) described the design process (Alexandrova 2006; Alexandrova and Northcott 2009). According to them, the theoreticians overstate their case when they claim that the FCC "chose an innovative form of auction . . . because theorists predicted it would induce more competitive bidding and a better match of licenses to firms" (McAfee and McMillan 1996, p. 160). Instead, they argue that game theory merely provided heuristics and pointed to problems that could possibly arise, and which experimentalists would then investigate by means of their testbed methodology. As their experiments delivered the bulk of the required evidence, the case in fact shows how limited the

theory alone is.[5] Generalizing this analysis, Alexandrova (2008) argues for a weak reading of models, according to which their role is mainly to suggest causal hypotheses about a target. Empirical studies are then needed to provide evidence concerning whether the causal hypothesis does indeed hold.

In sum, existing contributions have noted the limitations of theory for design purposes and have highlighted the need for supplementation with experiments. I will next introduce a different case study, which suggests additional roles for models and experiments in the service of design, as well as additional methods besides models and experiments in the design process.

3. The Matching Market for Medical Residents

Before becoming doctors, medical graduates in the US are required to take up training positions, or "residencies," in hospitals. These allow the "residents" to specialize in a specific medical branch. In general terms, the labor market for these positions works as follows: After interviews take place, the NRMP—a private, non-profit organization—collects rank order lists ("ROLs") from both applicants and hospitals. These lists reflect the applicants' preferences concerning the hospitals they had previously had an interview with, and the hospitals' preferences concerning the applicants they had interviewed. Assignments are then determined using a matching algorithm. Since the residencies shape the residents' future careers, and residents provide a significant source of the labor force for hospitals, it is vital that the market is organized fairly and efficiently.

In the early years after the NRMP's foundation in 1952, the matching procedure worked to the satisfaction of the participants, as indicated by participation rates in the system of over 95% (participants are free to decide whether to find matches through the centralized procedure or on their own). However, over the years several changes occurred. Initially, residents were predominantly male, and when female residents entered the market in the 1970s, there were increasing numbers of married couples who graduated from medical school together. Members of couples may have interrelated preferences, particularly to find positions close to one another. For example, even if a member of a couple individually prefers, say, a residency in Boston

[5]Nik-Khah (2008) goes further, claiming that "[g]ame theory was demonstrably irrelevant to most important aspects of the construction of markets, as evidenced by the need of the FCC to call on experimentalists" (p. 92). This interpretation cannot be sustained: even on a weak reading in which the *only* role of game-theoretical models was to provide heuristics and point to possible problems, this role alone makes theory clearly relevant for the design of the FCC auctions.

to one on the West Coast other things being equal, these preferences may switch if their partner attains a position in Los Angeles. In other words, couples give rise to complementarities, similar to those encountered in the FCC auctions. But the original matching algorithm could not accommodate complementarities because it only processed single preference lists. As a response to increasing discontent and declining rates of participation, the NRMP modified the system by permitting couples to hand in pairs of ROLs together and to specify a "leading member." The algorithm would then match the leading member first, followed by an editing of the other member's ROL to eliminate positions far from that of the leading member. However, this rather ad-hoc modification did not prevent participation rates from dropping.[6]

In the 1990s, the dissatisfaction among applicants—as expressed by various student associations—was at a peak. Many claimed that the algorithm was biased against them. Moreover, there was a rumor that applicants could "game the system" by submitting ROLs that didn't reflect their true preferences. Consequently, some student associations requested a change of the matching algorithm, or that the applicants be given more information on how to hand in their ROLs strategically.

The Board of Directors of the NRMP reacted in 1995, commissioning Alvin Roth to direct the design of a new matching procedure. They set three policy goals that should be achieved to the greatest degree possible: to incentivize applicants and hospitals to stick to the matchings (i.e., not to make arrangements outside the system); to make the matchings as favorable as possible for the applicants; and to reduce opportunities for strategic behavior. The new algorithm, which is now known as the "Roth-Peranson algorithm,"[7] was first introduced in 1998 and has been working successfully since.

In order to answer how Roth and his collaborators reformed the market, I will next describe in depth some of the models and other tools they used. I will first sketch a simple model of the market and some of the theoretical results that hold in this model. Then I will explain how the simple model was

[6]The accommodation of couples was not the only challenge the NRMP was facing. Hospitals may have interlinked numbers of positions such as, say, five in the neurology department if internal medicine fills all its positions, but fewer otherwise—a source of a different kind of complementarity (see Roth and Peranson 1999 for a description of all the complementarities present in the market). Furthermore, the numbers of graduates relative to residencies offered increased substantially over the years, which led to matchings being less favorable for the former.

[7]Elliott Peranson is founder of the National Matching Services Inc., a company devoted to providing matching solutions by implementing what they advertise as a "Nobel-Prize acclaimed algorithm" (https://natmatch.com/#, accessed on 09/05/2020).

manipulated and complemented with other tools to create an algorithm with desirable properties, and how the algorithm was eventually implemented.[8]

3.1. Modeling the Market

From a game theory perspective, the applicants' and the hospitals' preferences, together with a matching mechanism, define a game, in which their actions are to submit ROLs (or to opt out). Formally, there is a set of applicants $A = \{a_1, \ldots, a_m\}$ and a set of hospitals $H = \{h_1, \ldots, h_n\}$. Each hospital h_i offers a number of residencies which is specified by a quota, q_i. The preferences $\{\succ_{a_1}, \ldots, \succ_{a_m}, \succ_{h_1}, \ldots, \succ_{h_n}\}$ are assumed to be transitive, irreflexive, and complete lists for applicants of the hospitals they had an interview with and that they deem acceptable, and for hospitals of the applicants they had interviewed and whom they deem acceptable. Just like their preferences, the applicants' and hospitals' ROLs are transitive, irreflexive, complete lists of acceptable partners on the other side of the market. Note, however, that agents can be strategic, viz. submit ROLs that do not truthfully reflect their preferences.

A matching mechanism maps combinations of ROLs to matchings, which are the outcomes of the game. Formally, a matching μ is a subset of $A \times H$ such that any applicant appears in at most one pair (i.e., is either matched or unmatched) and each hospital h_i appears in at most q_i pairs (i.e., is either full or has empty places). Let's look at the mechanism in use when the NRMP directors commissioned the new design. As shown in Roth (1984), in our simple model it is equivalent to the *hospital-proposing deferred acceptance algorithm*:

- In the first step, each hospital "proposes"[9] to the highest-ranked applicants on its ROL, until its quota is filled. Each applicant tentatively "accepts" the highest-ranked proposer on their ROL, and rejects the other proposers.
- In the n-th step, each hospital subject to rejections in step $n-1$ proposes to the highest-ranked applicants to whom it has not previously proposed until its quota is filled. Each applicant tentatively accepts the highest-ranked hospital on their ROL among the proposers and the

[8]My account is based on Roth (1984, 2002, 2003, 2013, 2015, 2018), Roth and Sotomayor (1990), Roth and Peranson (1999) and Kojima et al. (2013).

[9]It is common to describe the algorithm using the predicates "propose" and "accept"/"reject." Of course, this refers not to the agents' behavior in a decentralized market but to the algorithm's processing of the ROLs.

hospital she tentatively accepted in the previous step, and rejects the others.

- The process is repeated until there are no more proposals, at which point the applicants are matched to the hospitals whose offers they are holding (or remain unmatched otherwise).

As Gale and Shapley (1962) show, in the simple model described above, this algorithm always finds *stable matchings*. A matching is stable with respect to the ROLs submitted if no one is matched to an unacceptable partner (i.e., a partner that does not appear on their ROL), and there is no *blocking pair*: a pair that consists of an applicant and a hospital that are not matched to each other but each is higher-ranked on the other's ROL than some partner assigned to them in the matching.[10]

How could this simple model inform the redesign of the medical match? An important function was to make the policy goals precise and to design algorithms implementing them within the model. Stability seemed to formalize the first-order goal to remove incentives for making deals outside the system, by removing blocking pairs that have these incentives. (I will give more nuance to this view below.) With respect to the other goals—to make the matchings favorable for the applicants and to reduce their opportunities for strategic behavior—different stable algorithms can be compared in the simple model. For instance, the *applicant-proposing deferred acceptance algorithm* (which is equivalent to the hospital-proposing algorithm with the roles of the applicants and the hospitals switched) produces stable matchings that are weakly preferred by all applicants to all other stable matchings with respect to their ROLs submitted, whereas the hospitals weakly prefer the matchings from the hospital-proposing algorithm to all other stable matchings. So the applicant-proposing algorithm might be expected to implement the goal of producing stable matchings which are as favorable as possible for applicants, thus offering advantages over the hospital-proposing algorithm that was in effect. Furthermore, the applicant-proposing algorithm makes it a dominant strategy for the applicants to submit their true preferences, whereas the hospital-proposing algorithm does not make it a dominant strategy for either side of the market to reveal their preferences (an asymmetry

[10]The intuition behind the proof that this algorithm implements stability is simple: under this procedure, no one can be matched to an unacceptable partner, and there can be no blocking pair because, if an applicant a_j is ranked higher on a hospital h_i's ROL than a student matched to it, h_i must have applied to a_j at some previous step and been rejected. Thus a_j must have ranked h_i lower than her actual match and so (h_i, a_j) is not a blocking pair.

stemming from the fact that hospitals take multiple students whereas students are assigned to a single hospital).

However, the simple model lacks relevant features of the market. As described above, there are couples among the applicants that are permitted to hand in ROLs specifying pairs of positions. Couples are absent in the model above, but they can be added to it. Thus, another important way in which this model was used was through intervention: features of the market could be added that were previously missing and their interplay with policy goals could then be investigated. For instance, some of the theorems described above do not generalize to models with couples; in particular, it cannot be guaranteed that stable matchings exist and thus there is no algorithm, like the above, that would always implement stable matchings. We will come back to this problem below.

3.2. Fitting the Prototype

As noted above, stability seemed to formalize the policy goal of incentivizing market participants to stick to their assigned partners. However, this is a hypothesis on the basis of the model alone, which relies on stringent common-knowledge assumptions about the agents' actions. These are not satisfied in the medical match, where neither applicants nor hospitals see the ROLs others. Was a lack of stability, introduced by the couples into the market, really the cause of the market failures?

In order to answer this question, empirical evidence was needed. There were regional matching markets for physicians and surgeons in Britain, which served as natural experiments. Of the eight markets investigated, six used unstable mechanisms, and only two had survived by the time the study was made. The two surviving markets used stable mechanisms and both were performing well. This provided evidence for the importance of stability. Of course, it was still possible that the survival or dissolution of the different markets was due to factors other than stability. In order to dispel this doubt, environments were created in laboratory experiments in which the only difference was the algorithm in use (Kagel and Roth 2000). The experiments reproduced the field results, thus confirming that stability is key for achieving the policy goal. The experiments were thus used to confirm the model results: within the model, properties were defined that could correspond to policy goals, and mechanisms designed that implement those properties; then natural and laboratory experiments that mirror the model provided evidence that these properties "work" in the real world and can be brought about by the mechanisms.

Where does this leave us in the design process? From models in combination with experiments, the conclusion could be drawn that stability was key. However, the models also showed that stability could not be guaranteed in the target market, where couples are present. However, when the NRMP directors commissioned a redesign of the matching process, it wouldn't suffice to point out this impossibility result: what was needed was a well-functioning algorithm, even if it could not always find stable matchings.

A simple deferred acceptance algorithm (modified to process couples' ROLs specifying pairs of positions) would not achieve this—which explains the fact that when couples entered the market in the 1960s, rates of participation dropped.[11] Roth and Peranson (1999) investigated a modified, student-proposing deferred acceptance algorithm that seeks to find stable matchings by detecting blocking pairs and repairing them, if possible, at intermediate steps. Because the set of stable matchings can be empty, there was of course no guarantee that the Roth-Peranson algorithm would always find a stable matching. In order to estimate the magnitude of this problem, they engaged in various *computational experiments*: runs of the algorithm using ROLs from previous years, as well as randomly generated ROLs. These experiments suggested that, under certain conditions (such as short ROLs and not too large a proportion of couples), stable matchings exist with a high probability in large markets. These conditions are fulfilled in the NRMP, where, for example, the ROLs are short because applicants interview at only a small fraction of the residencies. Being a large market, these results gave evidence that in the NRMP there is a high probability that the set of stable matchings is non-empty.

The computational experiments also suggested another important fact, namely that set of stable matchings, while non-empty, would be *small*. This is significant because, if the set is large, any stable matching algorithm will be biased towards some market participants, for instance in the way that the applicant-proposing deferred acceptance algorithm favors applicants over hospitals in a simple market without couples.[12] But when the set of stable

[11]Roughly, the problem is the following. Suppose an applicant-proposing deferred acceptance algorithm is running, and the members of a couple are both tentatively accepted by two programs. Then, if in the next step the first (but not the second) gets displaced by a preferred applicant, the couple applies to the next best preferred pair of positions which means that the second member of the couple is withdrawn from the hospital that had tentatively accepted her. But then blocking pairs may occur between that program and applicants it has rejected in order to hold the second couple member.

[12]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).

matchings is small, an algorithm producing stable matchings will be unbiased, as there are few applicants and hospitals that are matched differently under different stable matchings. Furthermore, there will be few opportunities for strategic behavior if the sets of stable matchings are small. Indeed, the Roth-Peranson algorithm practically makes it a dominant strategy for applicants and programs to state their true preferences.

The model results were thus complemented not only by experiments in the field and the lab, but also by computational experiments: model results located problems, and suggested computational experiments to investigate magnitudes that were, by the time, not known from the model. Interestingly, these experiments in turn prepared the ground for new theory. For instance, the computational experiments suggested that there might be theorems showing the existence of stable matchings in large markets with couples. This intuition turned out to be correct about a decade later, when it was proven analytically that, if there are sufficiently small numbers of couples and ROLs are short, as a market becomes large, the probability that a stable matching exists tends to certainty (Kojima et al. 2013).

3.3. Implementing the Design

As the new algorithm was found to exhibit desirable features—viz. that it would find stable matchings with a high probability, that it was unbiased, and that it left very little room for strategic behavior—Roth advised the directors of the NRMP to implement it. The implementation involved political issues, such as mediating between different stakeholders, in particular student associations and residency programs. The main concern was the following. As we saw earlier, the redesign of the medical match was commissioned as a response to severe market failures, damaging confidence in the NRMP on the part of the graduate applicants. When their distrust was at its peak, there was the impression that the market was biased against them. For this reason, one of the directors' policy goals was that the new algorithm should achieve stable matchings as favorable as possible for the applicants. But changing from a hospital-proposing algorithm to one that is essentially student-proposing might have conveyed the opposite impression, namely that the market would now be biased towards the applicants at the expense of residency programs. We know that this is not the case because, as the set of stable matchings is small, only a small number of applicants receive different matches at different stable matchings, and thus there is no room for systematic bias. However, for smooth implementation, the algorithm not only needed to exhibit this desirable feature; this also had to be conveyed to the market participants.

To this end, the designers made both the design process and the final result transparent. For instance, during the design process Roth posted progress reports on a web page and when the design was finalized, he presented the main results to various organizations of residency programs. The fact that the set of stable matchings was small convinced stakeholders that the algorithm wasn't systematically biased, in particular, that it would not favor applicants at the expense of residency programs. Consequently, the new algorithm did not face opposition and the NRMP directors decided to implement it. The algorithm is generally regarded as well-functioning and has been adopted in numerous labor markets around the world.

The NRMP continues to arouse economists' interest and there may be opportunities for further enhancements. The following are but three contemporary topics worth mentioning. First, empirical studies have been conducted to quantify the extent to which applicants rank programs truthfully (Rees-Jones 2018). Second, the application and interview processes that precede the matching procedure have increasingly been scrutinized (Echenique et al. 2020). During the COVID-19 pandemic, for instance, the likely impact of the pandemic on the interviews has been investigated: since virtual interviews might lead to excessive numbers of interviews being conducted, it has been proposed to cap the number of interviews a student can accept (Hammoud et al. 2020). And third, a culture has developed in which some hospital chairs exert pressure on the directors of residency programs to recruit their top-ranked applicants in the match (see Rozenshtein et al. 2020 for a survey from radiology). This may produce perverse incentives for program directors to rank applicants contrary to their true preferences or to put pressure on some applicants to rank their programs first, thus increasing their programs' performance (or perceived performance) in the matchings. These practices may introduce novel biases into the match and the NRMP disapproves of it. But even though there are ideas for how this issue might be alleviated (e.g., not allowing directors to share ROLs with their chairs), the problem remains to be fixed.

As these contemporary investigations show, the general focus has shifted from the algorithm design—which is deemed a success—to the broader environment in which the matching algorithm operates, which involves cultural factors that are not easily foreseeable.

4. Towards a General Account of Economic Design

The design processes of the spectrum auctions and of the medical match differed in various important ways. Most obviously, in the design of the medical match, a centralized matching system already existed, which had to be

reformed, whereas the spectrum auctions had to be designed from scratch. Partly for this reason, the relative importance of models, lab and field experiments, and computational experiments differed in the two cases. For instance, the history of the medical match, in combination with models, provided rich evidence of possible sources of market failure, in particular the lack of stability due to increasing numbers of couples. In contrast, in the auctions, where field data were largely absent and the models available more circumscribed, experimental test beds were heavily drawn upon. While acknowledging that different design efforts will generally differ from each other, Roth and Peranson argue that, "if we are to develop a body of knowledge about design practice in economics, we need to think about the methodological issues that may be common to many design efforts" (1999, p. 769). Focusing on common methodological issues from these two design processes, I will next offer an account of economic design that may serve as the starting point for the development of a structured body of knowledge about design practice, by integrating the three metaphors of the economist as designer, engineer and plumber.

4.1. The Designer

The economist-as-designer generally kicks off the design process by constructing and manipulating models. In both cases that have been considered here, models were available from previous theory: in the medical match previous matching theory and in the spectrum auctions previous auction theory provided the basic models. These were abstract, theoretical models with only very loose connections to real-world institutions. The models thus neither provided accurate representations of existing markets, nor of the prospective markets that would eventually be implemented. Rather, they formalized basic market structures, such as a generic auction or matching market. Subsequently, the designers manipulated these models in order to learn more about their specific target of interest.

The practice of taking abstract theoretical models and manipulating them is common practice in economics (Morgan 2012); but the economist-as-designer faces a peculiar challenge. In more typical (non-design) modeling practices, the relevant target provides important constraints on how a model ought to be constructed and manipulated: the model can make some idealized assumptions, such as agents' perfect rationality and computational abilities, but it is not the case that anything goes. For example, if one were to model a certain market but came up with a model of a completely different market structure than that of the intended

market, the result would be considered a modeling failure.[13] But in the case of economic design, the designer cannot simply model a target, because she aims to change, or create, the target itself. Thus, the question is, how can the target provide constraints on how the model ought to be constructed and manipulated if the target is itself a counterfactual possibility? What guides designers in their modeling practices?

The answer is that designers are guided not only by positive but also by *normative constraints*, that is, by the policy goals stipulating for a given case what kinds of outcomes a design ought to achieve (e.g., incentivizing applicants and hospitals to stick to matchings in the case of the medical match, efficiency in the case of the spectrum auctions).[14] These normative constraints take the place of missing positive constraints in the designer's construction and manipulation of models. To make this more precise, we must distinguish between the "moving" and "fixed" parts of the target system. There are parts of the target, such as people's preferences, that cannot easily be changed and which the designer thus assumes to be fixed. These impose the positive constraints on the model, defining what may be called the "environment" (Hurwicz 1973). In contrast, the moving part for the designer is the mechanism, that is, the algorithm mapping possible combinations of actions into outputs. For instance, in the case of the medical match, the mechanism was a matching algorithm, which had to be redesigned in such a way that would bring about set policy goals when implemented in the real world. Thus, while the environment formalizes the positive constraints, the designed mechanism must implement the normative constraints for a design to be successful.

Hence, the designer seeks to build a mechanism implementing the normative constraints while respecting the environment defined by positive constraints. The designer's task is distinctive as it integrates positive and normative modeling practices in this specific way. She generally proceeds by manipulating the initial models, in order to understand how different positive and normative constraints interact. Let's see how this works in practice. For example, in the medical match, the goal of removing incentives for making deals outside the system was formalized as stability in a simple model. Within this model, mechanisms could then be designed that implement this goal. As we saw earlier, the initial model ignored important positive constraints, in

[13]My discussion does not depend on the difficult question of what should count as modeling failures (on this question see Mäki 2011), but only on the assumption that there are clear cases of such failures.

[14]We made the assumption of exogenous policy goals for simplicity: other factors, including ethical considerations, may codetermine what is seen as normative constraints in a given case.

particular applicants' preferences to be matched to positions near their partners. These were not satisfied in the environment of this model, where couples were missing, but the model could be manipulated by adding them. However, by adding couples to the model, it could be shown that stable matchings might not exist, hence, in these cases, no mechanism could achieve them. Thus, by manipulating the model, it was discovered that it may be impossible to satisfy an important normative constraint, namely stability, under some positive constraints that will obtain in the real market.

In this example, positive constraints inhibited the achievement of certain normative constraints in the model, but there are other possibilities. For instance, formalizing policy goals within the model will also make it plain when a combination of goals cannot jointly be satisfied, in which case the model provides helpful feedback to the policy maker setting those goals (Li 2017). By manipulating models to see how different constraints interact with each other, the designer will also learn the limits of what can possibly be implemented.

Summing up, the designer typically takes abstract theoretical models from existing theory, which formalize a generic market. Subsequently, she manipulates these models to satisfy important positive and normative constraints and to understand how these constraints interact. In this process, the models serve various important purposes. First, they are used to make policy goals precise by formalizing them as normative constraints. We will see below that, whether a formalization is "correct" will subsequently be tested empirically; but without a model, it might not be known what to strive for in the first place. Second, a prototype mechanism, or class of such mechanisms, can be developed, which implement those goals within the model. Models thus provide guidance concerning what kinds of mechanisms may be worth testing further, excluding those mechanisms that have no chance of producing desirable properties. Narrowing down the potentially infinite number of mechanisms to those that could possibly lead to the implementation of these properties is important as this may be impossible, and certainly inefficient, through trial and error. And finally, the model allows us to discover how constraints interact and the limits on what can possibly be implemented. These interactions and limits can subsequently be investigated quantitatively. Here, the engineer enters the stage.

4.2. The Engineer

The economist-as-engineer inherits the designer's prototype mechanism and fits it to the real world. The engineer's main tools are experiments and

computation, combined with the designer's models. Let's consider the use of these tools in turn.

As philosophers of science have noted, lab and field experiments can serve a variety of important purposes (Guala 2005). One such purpose is in testing whether a model result holds water. Models make false assumptions and omissions as the inclusion of all real-world constraints would make them intractable. Thus, their results must be tested in the lab or field. For example, the properties formalizing policy goals are defined relative to the assumptions of the model, such as idealized preference structures, restricted strategy sets, or full rationality, which may not obtain in the real world. Do these properties (e.g., stability) correspond to what the policy maker wants to achieve with set goals? And will the suggested mechanisms work in achieving these goals outside the model? In the redesign of the medical match, field and lab experiments provided evidence that stability went a long way towards implementing the policy goals and that in the absence of match variations such as couples, student-proposing deferred acceptance algorithms are generally well-behaved in achieving them. But the applications of experiments are richer than merely testing model results; experiments can also be used to discover facts not previously known from theory. In the design of the spectrum auctions, experimental testbeds were decisive in choosing between different auction formats where theory was silent. Moreover, experiments were used to develop the chosen format further, for instance, by providing evidence that package bidding auctions can improve efficiency in the presence of complementarities.

While philosophers of science have investigated the use of lab experiments in the design of the FCC auctions, the use of computation has not received much attention. But just as in the NRMP case, computational experiments played an important role in the design of the FCC auctions. In particular, they were crucial for the FCC's decision in the mid-2000s to increasingly conduct package bidding auctions. Let's have a look at how computational experiments were used in both cases.

There are interesting structural similarities between package bidding auctions and matching markets with couples (Roth and Sotomayor 1990). In particular, they deal with complementarities analogously: in a matching market with couples by permitting couples to submit complementary preferences, and in package bidding auctions by permitting bidders to bid for packages of licenses. We have seen that these complementarities led to problems in both cases, and that computational experiments were important in dealing with these problems. In the case of the medical match, the problem was that stable matchings might not exist when couples are present in the market; while in the spectrum auction case, the worry was that it might be

intractable for the auctioneer to select the winning bids because of the large number of possible packages. In both cases, the engineers made use of computational experiments to investigate the magnitude of these problems. In the medical match, computational experiments suggested that, under certain conditions, including not too large a proportion of couples, as the market becomes large the set of stable matchings is unlikely to be empty. As these conditions were fulfilled in the actual market, there was hope that stable matchings might be found. Similarly, in the spectrum auction case, computational experiments showed that selecting winning bids is not intractable if the number of packages bidders are allowed to bid on is restricted (see De Vries and Vohra 2003), consequently the FCC restricted this number to 12 packages when initially moving to package bidding auctions. Computation thus played a similar role in both cases, suggesting ways around problems, which in both cases involved limiting the complementarities in the markets in some way (where these limits were either satisfied naturally or could be imposed by the design).

The inherited models from the designer have important roles to play for the engineer. They suggest hypotheses—about the meaning of policy goals, mechanisms, or trade-offs between constraints—which can then be tested in lab, field or computational experiments. But this does not exhaust their role, as the converse may happen too: sometimes experiments suggest hypotheses, which may in turn be confirmed analytically in the models. This was most conspicuous in the case of the medical match, where, as we have seen, the large market hypothesis, suggested by computational experiments, was later analytically proven to be correct. Thus, while the designer's models typically "come first," both chronologically and epistemically, there may be feedback between the models and the engineer's experiments: not only do model results lead to experiments, but the converse is also true. This lesson is also true of the spectrum auctions, where lab experiments sparked a comprehensive body of new theory. It is in this two-way sense that, for the engineer, experiments and models complement each other (Roth 2002). Once the engineer has fitted the prototype to the target through the application of experiments in combination with models, the plumber enters the stage.

4.3. The Plumber

Duflo (2017) defines plumbing as an implementation strategy: the economist-as-plumber installs the mechanism that the designers and the engineers have settled upon in the real world, observes the effects and mends the design if problems emerge. As the name suggests, the plumber needs to take a more

tentative approach than the engineer, as issues may arise that neither theory nor experiments could quantify or even foresee. We encountered such unforeseen issues in the redesign of the medical match, but they also occurred in the design of the spectrum auctions. Cramton and Schwartz (2000) investigate a variety of unforeseen ways in which bidders bid collusively in the early FCC auctions, and they propose possible solutions. For instance, they observe that some bidders engaged in "code bidding": a cunning way of bidding collusively, in which bidders use the last digits of their bids in order to signal license numbers to other bidders. (With bids of six digits or more, the cost of signaling license numbers, which are no more than three digits, are negligible.) Bidders used the signals, for example, to make it plain to other bidders that they would punish them if they were to bid on the signaled license, thus increasing their own chances to win the license at a low price. Because such collusive behavior may affect efficiency, various rule changes were issued after it was observed, for instance, bids were restricted to fixed increments on standing bids, which mitigated code bidding.

Thus, the plumber must find solutions to unforeseen issues when the machine already is in motion. Plumbing is not restricted to fixing issues by mending the mechanism in use: often, plumbers must keep an eye on details of the broader environment in which the mechanism operates because cultural factors, political processes and attitudes of market participants towards the design may affect its functioning. For instance, it was crucial in the case of the medical match that market participants' trust in the system could be restored when the new design was implemented, by convincing them that the new algorithm wasn't biased against them and that they could safely reveal their true preferences. As we have seen, this was achieved by transparently communicating that the set of stable matchings was small and hence, that there was no room for systematic bias.

By taking into account "soft" factors such as trust and transparency, the plumber conceives of individuals less as unboundedly rational maximizers the way that designers typically do. Of course, insofar as the plumber inherits the broad mechanism from the prior design and engineering processes, she will rely on the rationality assumptions of the designers and the engineers when installing the mechanism in the target, as the resulting institution will work best if these assumptions are approximately true. But as unforeseen issues are detected and the mechanism changed in response, the plumber might strive to make the mechanism increasingly independent of stringent rationality assumptions. Thus, by installing mechanisms and fixing them as problems arise, the plumber also explores the extent to which assumptions from previous theory and experiments hold in the real world, and what the limits of their results are.

The exploration of prior theory and experiment distinguishes plumbing as it is applied in the service of economic design from a "plumbing-only" approach, in which the plumber would not resort to previous results. For instance, proponents of plumbing-only might conduct randomized controlled trials (RCTs) to motivate a certain policy, as RCTs require only minimal theoretical assumptions. The increasing use of RCTs in this way has been criticized because less can be learned from RCTs alone about relevant causal mechanisms and thus extrapolation to the relevant target may be difficult (Deaton and Cartwright 2018). But plumbing applied for economic design is not plumbing-only because the designer's and the engineer's results are built upon and enhanced.

4.4. Summing Up

Economic design can be seen as a three-stage process, wherein the economist adopts a specific role at each stage:

First, the economist-as-designer creates models in which policy-goals can be made precise, and mechanisms that could possibly bring those goals about are constructed. The designer proceeds by combining positive and normative modeling practices: defining an environment that formalizes important positive constraints and, within this environment, mechanisms that may satisfy normative constraints. By manipulating the model, trade-offs between these constraints, and limits to what can possibly be implemented, are explored.

Second, the economist-as-engineer conducts lab, field and computational experiments in order to diversify evidence and to investigate previously identified trade-offs and limits quantitatively, which may provide ways out of impossibility results and may allow the engineer to refine the algorithm. In this process, the engineer's experiments and the designer's models are used complementarily and there may be feedback in both directions.

Third, the economist-as-plumber implements the design in the real world and mends it as problems arise. The plumber pays attention not only to the algorithm itself but also the environment in which it operates, as cultural and political factors—such as whether market participants believe that the designed algorithm is biased against them or trust that the designed algorithm is unbiased—can contribute to the failure or functioning of a design.

I should hasten to add a caveat: the three stages of a design process are not always perfectly clearly distinguishable and often, such as in the case of the medical match, one and the same economist will take on the role of the designer, the engineer and the plumber. Furthermore, while the order of the stages is most conveniently presented as, "design followed by engineering followed by plumbing," a given design process may not be as ordered, as

for instance new theory may be generated as a response to the engineer's experiments, or experiments conducted to understand issues that the plumber observed. That being said, the designer, the engineer and the plumber clearly adopt crucial roles in the cases of economic design discussed here, and their roles are sufficiently separable for the distinction between these roles to provide a useful heuristic for clarifying and systematizing the design processes. In line with the argument that we need to find common denominators of many design cases in order to develop a body of knowledge (Roth and Peranson 1999), I suggest that this heuristic is a useful starting point for a general account of economic design. On the way to a general account, further cases of economic design should be analyzed and the heuristic refined; but a general account of economic design should certainly be in line with the two flagship cases discussed here, which underwrite my proposed distinction.

5. Concluding Remarks

I have argued that, when designing an institution, economists wear the hats of the designer, the engineer and the plumber, respectively. My account augments existing studies in the philosophy of science that have exclusively focused on spectrum auctions. By considering the redesign of the medical match, and exploring its commonalities with these auctions, various important and hitherto unnoticed methodological features have been uncovered. These include the combination of positive and normative modeling in the design stage, which has been made precise; the important role of computation, both for epistemic goals and in algorithm design in the engineering stage; and the importance of the environment within which a mechanism is implemented in the plumbing stage.

The account can be applied to classify and distinguish full-blown economic design from more sparse practices, for instance, a plumbing-only approach, namely by checking whether each stage is present in the case at hand. But the account can also be given a normative reading. A recurrent theme has been that the designer's, the engineer's and the plumber's roles in building an institution complement each other in important ways. This suggests that doing with less than the full designer-engineer-plumber line-up may fall short of generating a well-functioning institution. Indeed, a plumbing-only approach might fail to discover relevant causal mechanisms that must be understood in order to design an institution that does what it is supposed to. Similarly, "design-only"—proposing the implementation of broad mechanisms for specific social reforms—would be unlikely to yield a reliable institution as it would pay insufficient attention to the fine-grained

details of the specific target (see Levine 2020); but these details, for instance differences in people's skills in strategizing, may result in undesirable and unforeseen consequences, such as biases, if they are not attended sufficiently. Want to build a well-functioning institution? Engage a designer, an engineer, and a plumber.

Acknowledgements

For helpful discussions and comments at various stages of this research, I would like to thank Anna Alexandrova, Luc Bovens, Donal Khosrowi Djen-Gheschlaghi, Bryan Roberts, Lucie White, two anonymous referees of this journal, and the audiences at the following conferences and workshops: the Soul of Economics Conference, University of Zurich; the Network for Economic Philosophy Kick-Off meeting, Aix-Marseille School of Economics (AMSE); and the Philosophy of Social Science Roundtable, Emory University, Atlanta.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/ or publication of this article: This research was funded in part by the Volkswagen Foundation within the project 'Bias and Discrimination in Big Data and Algorithmic Processing: Philosophical Assessments, Legal Dimensions, and Technical Solutions.'

References

Alexandrova, Anna. 2006. "Connecting Economic Models to the Real World: Game Theory and the FCC Spectrum Auctions." *Philosophy of the Social Sciences* 36: 173-92.

Alexandrova, Anna. 2008. "Making Models Count." *Philosophy of Science* 75: 383-404.

Alexandrova, Anna, and Robert Northcott. 2009. "Progress in Economics: Lessons from the Spectrum Auctions." In *The Oxford Handbook of Philosophy of Economics*, edited by Harold Kincaid and Don Ross. Oxford: Oxford University Press.

Ausubel, Lawrence M. and Paul Milgrom. Ascending auctions with package bidding. The B.E. *Journal of Theoretical Economics* 1(1):1-44, 2002.

Cartwright, Nancy. 2009. "If No Capacities Then No Credible Worlds. But Can Models Reveal Capacities?" *Erkenntnis* 70(1): 45-58.

Committee for the Prize in Economic Sciences in Memory of Alfred Nobel. Scientific Background on the Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2020: Improvements to Auction Theory and

Inventions of New Auction Formats. The Royal Swedish Academy of Sciences, 2020. https://www.nobelprize.org/uploads/2020/09/advanced-economicsciencesprize2020.pdf, Accessed on 18/10/2020.

Cramton, Peter, and Jesse A. Schwartz. 2000. "Collusive Bidding: Lessons from the FCC Spectrum Auctions." *Journal of Regulatory Economics* 17(3): 229-52.

Deaton, Angus, and Nancy Cartwright. 2018. "Understanding and Misunderstanding Randomized Controlled Trials." *Social Science & Medicine* 210: 2-21.

De Vries, Sven, and Rakesh V. Vohra. 2003. "Combinatorial Auctions: A Survey." *INFORMS Journal on Computing* 15(3): 284-309.

Duflo, Esther. 2017. "The Economist as Plumber." *American Economic Review* 107(5): 1-26.

Echenique, Federico, Ruy Gonzalez, Alistair Wilson, and Leeat Yariv. 2020. "Top of the Batch: Interviews and the Match." arXiv:2002.05323. https://arxiv.org/abs/2002.05323

Gale, David, and Lloyd S. Shapley. 1962. "College Admissions and the Stability of Marriage." *American Mathematical Monthly* 69(1): 9-15.

Guala, Francesco. 2001. "Building Economic Machines: The FCC Auctions." *Studies in History and Philosophy of Science* 32(3): 453-77.

Guala, Francesco. 2005. *The Methodology of Experimental Economics*. Cambridge: Cambridge University Press.

Guala, Francesco. 2006. "Getting the FCC Auctions Straight: A Reply to Nik-Khah." *Economic Sociology the European Electronic Newsletter* 7(3): 23-8.

Guala, Francesco. 2007. "How to Do Things with Experimental Economics." In *Do Economists Make Markets?*, edited by D. MacKenzie, F. Muniesa, and L. Siu, 128-62. Princeton: Princeton University Press.

Hammoud, Maya M., Taylor Standiford, and J. Bryan Carmody. 2020. "Potential Implications of COVID-19 for the 2020-2021 Residency Application Cycle." *The Journal of the American Medical Association* 324(1): 29-30.

Hurwicz, Leonid. 1973. "The Design of Mechanisms for Resource Allocation." *The American Economic Review* 63(2): 1-30.

Kagel, John H., and Alvin E. Roth. 2000. "The Dynamics of Reorganization in Matching Markets: A Laboratory Experiment Motivated by a Natural Experiment." *The Quarterly Journal of Economics* 115(1): 201-35.

Knuth, Donald E. 1996. *Stable Marriage and Its Relation to Other Combinatorial Problems. CRM Proceedings and Lecture Notes*. Vol. 10. Providence: American Mathematical Society.

Kojima, Fuhito, Parag A. Pathak, and Alvin E. Roth. 2013. "Matching with Couples: Stability and Incentives in Large Markets." *The Quarterly Journal of Economics* 128(4): 1585-1632.

Levine, David K. 2020. "*Radical Markets* by Eric Posner and E. Glen Weyl: A Review Essay." *Journal of Economic Literature* 58(2): 471-87.

Li, Shengwu. 2017. "Ethics and Market Design." *Oxford Review of Economic Policy* 33(4): 705-20.

Mäki, Uskali. 2011. "Models and the Locus of Their Truth." *Synthese* 180(1): 47-63.

McAfee, R. Preston, and John McMillan. 1996. "Analyzing the Airwaves Auction." *Journal of Economic Perspectives* 10(1): 159-75.

Morgan, Mary S. 2012. *The World in the Model: How Economists Work and Think*. Cambridge: Cambridge University Press.

Nik-Khah, Edward. 2008. "A Tale of Two Auctions." *Journal of Institutional Economics* 4(1): 73-97.

Rees-Jones, Alex. 2018. "Suboptimal Behavior in Strategy-Proof Mechanisms: Evidence from the Residency Match." *Games and Economic Behavior* 108: 317-30.

Roth, Alvin E. 1984. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy* 92: 991-1016.

Roth, Alvin E. 2002. "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics." *Econometrica* 70(4): 1341-78.

Roth, Alvin E. 2003. "The Origins, History, and Design of the Resident Match." *Journal of the American Medical Association* 289: 909-12.

Roth, Alvin E. 2013. "What Have We Learned From Market Design?" In *The Handbook of Market Design*, edited by N. Vulkan, A. E. Roth, and Z. Neeman. Oxford: Oxford University Press.

Roth, Alvin E. 2015. *Who Gets What – And Why*. Boston: Houghton Mifflin Harcourt.

Roth, Alvin E. 2018. "Marketplaces, Markets, and Market Design." *American Economic Review* 108(7): 1609-58.

Roth, Alvin E., and Elliott Peranson. 1999. "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design." *American Economic Review* 89(4): 748-80.

Roth, Alvin E., and Marilda A. Sotomayor. 1990. *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge: Cambridge University Press.

Rozenshtein, Anna, Brent D. Griffith, Priscilla J. Slanetz, Carolynn M. DeBenedectis, Jennifer E. Gould, Jennifer R. Kohr, Tan-Lucien Mohammed, et al. 2020. ""What Program Directors Think" V: Results of the 2019 Spring Survey of the Association of Program Directors in Radiology (APDR)." *Academic Radiology*. Published electronically August 8 2020. doi:10.1016/j.acra.2020.06.035.

Author Biography

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