Ethics and Market Design

Li, Shengwu

Harvard Society of Fellows

25 July 2017

Ethics and Market Design

Shengwu Li[*]

July 25, 2017

1 Introduction

> [...] just as there is a chemical engineering literature (and not just literature about theoretical and laboratory chemistry) and a medical literature (and not just a biology literature), economists need to develop a scientific literature concerned with practical problems of design. (Roth and Peranson, 1999)

> It is, in fact, arguable that economics has had two rather different origins [...] concerned respectively with 'ethics' on the one hand, and with what may be called 'engineering' on the other. (Sen, 1987)

The word "market" often refers to a set of agents who seek to transact with each other. In this sense, the stock market consists of individuals and corporations seeking to buy and sell stocks, and the market for medical residents consists of doctors and hospitals. However, the word "market" can also refer to the rule-governed institution that facilitates these transactions. In this sense, a stock market is a platform, such as the New York Stock Exchange, that enables participants to find willing trading partners according to a set of common rules.[1] Similarly, the market for medical residents in the United States consists of a clearinghouse, the National Resident Matching Program, that matches doctors to hospitals using a centralized algorithm that accounts for the preferences of both sides (Roth and Peranson, 1999).

An institution that facilitates transactions can be designed - the rules affect who trades, what is traded, and the terms of those trades. Market design takes an engineering approach to the problem: It seeks to construct rules that facilitate trade and lead to good outcomes. Market designers tend to adopt fairly expansive definitions of what counts as a transaction and (thus) what constitutes a market. They have investigated rules for allocating students to schools and for Internet dating platforms, even though laypeople may not see these as involving transactions at all (Abdulkadiroglu and Sönmez, 2003; Lee and Niederle, 2015).

---

[1]For instance, the rules of the New York Stock Exchange constrain the hours of trading, prohibit insider trading, and specify how buy and sell orders are processed in real time (Budish et al., 2015).

Plainly, in the course of studying market design we encounter ethical questions: Should public schools give priority to students living nearby, even if this entrenches existing economic inequality? Should wireless spectrum licenses have a limited lease or persist indefinitely, and does it matter what future generations would want? Should adults be permitted to trade kidneys for money?

The central question of this paper is: *How (if at all) should economic engineers think about ethics?* If we are to evaluate a market design, we require normative criteria. Standard welfare analysis offers a preference utilitarian framework; only individual welfare matters, welfare is equivalent to preference satisfaction, and preference satisfaction should be maximized (in some aggregate sense).[2] However, it is not clear that preference utilitarianism is the correct ethical position, and many stakeholders and policymakers are not preference utilitarians. Moreover, insightful papers on market design often consider non-utilitarian criteria such as fairness and privacy-preservation, which need not be grounded narrowly in preference satisfaction. The first thesis of this paper is that *the literature on market design does not, and should not, rely exclusively on preference utilitarianism to evaluate designs.*

If market design does not rely solely on preference utilitarianism, one is tempted to insist that market design should sort out its normative foundations before proceeding. We could seek an ethical position to replace preference utilitarianism, one that integrates concerns about fairness, rights, and individual agency that preference utilitarianism does not easily capture. Once market designers have a broad consensus on the correct normative criteria, we can then apply those criteria to evaluate markets.

I will argue that we should resist this temptation. There is the usual argument from specialization and comparative advantage; it is unclear that economists *qua* economists have any special ability to resolve ethical disagreements. Additionally, it is unclear when (if ever) this foundation-building project will succeed, so it is not reasonable to wait for its resolution. Furthermore, market design can produce valuable insights without relying on a comprehensive ethical foundation. The second thesis of this paper is that *market designers should study the connection between designs and consequences, and should not attempt to resolve fundamental ethical questions.*

I am not advocating a divorce between ethics and engineering. We should resist the opposite temptation, which is to avoid considering ethical questions at all, and to pretend that market design is a purely technical exercise with no normative content. Market design needs value judgments: We seek to design markets that function well, not badly, and "well" and "badly" have normative content. If market design has no explicit ethical criteria, it will rely on implicit criteria - often criteria selected for mere familiarity rather than truth.

The third thesis of this paper is that *the theory and practice of market design should maintain an informed neutrality between reasonable ethical positions.* We may not resolve the overarching ethical question, "How, all things considered, should one design a market?" But we can usefully formalize many smaller value judgments, such as whether a market is fair, is transparent, increases

---

[2] "The standard propositions of modern welfare economics depend on combining self-seeking behaviour, on the one hand, and judging social achievement by some utility-based criterion on the other." (Sen, 1987)

welfare, and protects individual agency. These we could regard as intermediate inputs to the ethical production function; they do not (without a complete ethical theory) settle the overarching question, but many answers to the overarching question take these variables as arguments. By investigating these variables, we can guide policymakers in designing markets, while remaining neutral between reasonable ethical positions. Market design already recognizes plural values. In sketching this framework, I aim to justify these aspects of the *status quo* and call for an expansion of current practice.

Since this paper is at the intersection of economics and philosophy, I will adopt a hybrid strategy to exposit the material. I will propose a series of simple models about the relationship between ethics and market design, but will discuss these models in prose rather than theorems. If the models are wrong (as all models are), they will at least be precise and transparent. We start with a *too simple* model, that (for several good reasons) is the default normative position in economics. We will add complications as needed.

2 A (Too) Simple Model of Ethics and Market Design

Consider any market governed by rules, from spectrum auctions, to school choice mechanisms, to kidney exchanges. That market has a set of feasible designs, $\mathcal{D}$, and a set of consequences, $\mathcal{C}$. Each design leads to some consequence, specified by a function $f : \mathcal{D} \to \mathcal{C}$.

I use the word 'consequences' to distinguish these from 'outcomes', which have a technical meaning in market design. Market design treats *outcomes* and *preferences* as distinct. An outcome specifies some partial state of affairs - an allocation of objects to agents, or a set of prices and quantities transacted, *etc.* These are treated as distinct from agents' preferences, which are defined over outcomes. By contrast, a *consequence* is a full state of affairs. It specifies not only the outcome, but also how that outcome ranks in agents' preferences. It even specifies features of the world that are affected by the choice of design, whether or not those are already captured by agents' preference satisfaction. For instance, the consequences of a market for carbon permits include the temperature of the earth in two thousand years' time. The consequences of a wireless spectrum auction include a price for mobile telephony services. The consequences of a kidney exchange include health measures such as quality of life and mortality, for both participants and non-participants. A consequence is richer than an outcome; it is a description of the effects of the market on the world.

In many market design problems, agents know some things that the designer does not. A company buying radio spectrum may know more than the auctioneer about next year's demand for wireless broadband. A doctor applying for medical residencies may know more than the clearinghouse about his preferences over hospitals. It is useful (and standard) to conceive both private information and private preferences as an agent's *type*, and to think of a design as determining a function from types to outcomes. In that case, the consequences of a design include this function - they specify what happens, contingent on the agents' types. If the designer has a prior belief about agent types, then the consequences of a design include a joint distribution of types and outcomes.

Now we build ethics into the model. For any consequence, we may be called on to make a value judgment. Is that consequence desirable? Should we bring it about? Are we morally *permitted* to bring it about? Are we morally *obliged* to bring it about? That is, there is some set of overall normative judgments $\mathcal{J}$, and a value function $v : \mathcal{C} \to \mathcal{J}$. We could then, with some abuse of notation, make value judgments about designs; that is, we could define $v(d) := v(f(d))$ for $d \in \mathcal{D}$. In this model, value judgments are *consequentialist*, in the sense that $v(d)$ depends only on $f(d)$; value judgments about a design are just equal to value judgments about its consequences.

The heart of the too simple model is this: Studying f is a problem for market design. Studying v is a problem for ethics. Given f and v, we can compute $v(f(d))$, and that settles the relevant ethical questions.

Standard welfare analysis supplies an answer for v; it specifies that v is *preference utilitarianism*.[3] This means that, when evaluating any consequence, only the welfare of agents matters.[4] On the preference utilitarian account, an agent's welfare is equivalent to that agent's preference satisfaction. This leads to an ordering $\succeq$ over outcomes, where $c \succeq c'$ if agents' preferences at least as satisfied in c as in c' (in some aggregate sense, that also accounts for uncertainty about types). If we are wary of interpersonal utility comparisons, then $\succeq$ may be a partial ordering using the Pareto criterion. However, in many settings the standard approach constructs a total ordering via a normalization. For instance, in the auctions literature the marginal satisfaction from one dollar of consumption is usually normalized to be equal across agents[5]. Preference utilitarianism then specifies that we are to bring about some best feasible consequence; that is, we are obliged to bring about $c \in f(\mathcal{D})$ such that there does not exist $c' \in f(\mathcal{D})$ such that: $c' \succeq c$ and not $c \succeq c'$. Statements about whether some design d is desirable, good, and so on depend on the position of $f(d)$ in the ordering.[6]

Preference utilitarianism is the default normative position in economics, mostly for good reasons. It is an intuitively plausible way to structure policy advice; human welfare is almost always an important policy objective, and often each agent knows what promotes her welfare better than we do. Preference utilitarianism is parsimonious. Agents' preferences over outcomes are already in (almost) every positive economic model, so we can evaluate a preference utilitarian v using only the model primitives that we used to investigate f. Furthermore, preference utilitarianism is tractable. It is easy to formalize in most settings of interest, and one can set about searching for the design that maximizes social welfare, a well-defined objective. A full adherence to preference utilitarianism would yield useful methodological unity; market designers could focus our efforts on investigating f through theory, observation, and experiments, and avoid normative disagreements about v that

---

[3]Sen (1987, Ch. 2) exposits and discusses the standard approach.

[4]A full theory would specify the morally relevant set of agents. When a design affects future generations (as in a market for carbon permits) or when a design affects which agents exist (as in a market for pregnancy surrogates), this may not be straightforward. Parfit (1984) discusses ethical problems that affect which agents exist.

[5]See Krishna (2010) for an exposition of the standard approach.

[6]It may at first seem that we are ruling out a cardinal interpretation of preference satisfaction, by specifying a binary relation $\succeq$. However, if we enrich $\mathcal{D}$ to include lotteries over deterministic designs and $\succeq$ is a total ordering satisfying the axioms of Von Neumann and Morgenstern (1953), then we can construct a cardinal interpretation using the usual method.

are difficult to settle with the methods of economics.

It is beyond the scope of this paper to argue that preference utilitarianism is false. For now, it suffices to argue that preference utilitarianism is not obviously true.

If preference utilitarianism is obviously the correct ethical theory, then one would expect a high degree of consensus among informed experts. However, a majority of professional philosophers do not believe that preference utilitarianism is true. In a survey of regular faculty members at 99 leading philosophy departments, only 24% of respondents reported believing in consequentialism (for comparison, 26% reported believing in deontology) (Bourget and Chalmers, 2014). Since preference utilitarianism is a species of consequentialism, this implies that no more than 24% of respondents believe that preference utilitarianism is the correct ethical theory. This is not strong evidence that preference utilitarianism is false, but it is evidence that preference utilitarianism is not obviously true.

There are important ethical dimensions of markets that preference utilitarianism does not capture. For instance, a market may be noxious because some of the transacting parties are vulnerable; "when a desperately poor person agrees to part with an asset at a fire sale price, even if the exchange improves his well-being we are rightly concerned with the fact that his circumstances made him willing to accept an offer for his asset that no one with a decent alternative would ever accept" (Satz, 2010, p. 97). It is difficult to express, in preference utilitarian terms, why a market that exploits well-informed but vulnerable people is ethically suspect. The very fact that a vulnerable person agrees to an exploitative transaction indicates that the transaction is a Pareto improvement. A preference utilitarian could regret the background circumstances that lead agents to be vulnerable. However, preference utilitarianism has difficulty with the intuitive claim that it is wrong to take advantage of another person's bad circumstances to make a profit, especially when this involves driving a hard bargain that he would not otherwise accept. It has even more trouble explaining why building markets that enable many such transactions could be ethically problematic.

Another ethical dimension of markets is that they can change the social meaning of the goods being traded. New York City's Public Theater puts on free Shakespeare performances in Central Park, but some attendees skip the queue by hiring others to queue on their behalf. When rich attendees can pay for priority at this public event, this undermines the event's role as "a kind of civic celebration", "a gift the city gives itself" (Sandel, 2012, p. 33). To take an older example, in Catholic doctrine an *indulgence* is "a remission before God of the temporal punishment due to sins whose guilt has already been forgiven"; it reduces the punishment that the believer must endure after death before entering heaven (Catholic Church, 2000). In the Middle Ages, church officials sold indulgences to believers in order to fund expensive projects such as crusades and cathedrals (Cross and Livingstone, 2005). This is now widely regarded as an abuse of the Church's authority; the sale of indulgences was condemned[7] in Martin Luther's Ninety-Five Theses, the document that sparked the Protestant Reformation. When Pope Leo X sold indulgences to reconstruct St.

---

[7]Theses 26 and 27 read, "They "preach human opinions" who say that as soon as a coin thrown into the money chest clinks, a soul flies [out of purgatory]. It is certain that when a coin clinks in the money chest profits and avarice may well be increased, but the intercession of the church rests on God's choice alone." (Wengert, 2015)

Peter's Basilica, he would have been ill-advised by a market designer arguing that the exchange of indulgences for money was a Pareto improvement; the sale of a sacred ritual altered its social meaning.[8]

Preference utilitarianism can capture some aspects of social meaning, but only in an expanded model, in which agents have preferences not only over allocations but also over social norms, and which specifies how different market designs lead to different norms. This expansion of the model makes preference utilitarianism less parsimonious and tractable, which diminishes its appeal as a normative framework. When engaging in market design, it may often be useful to treat concerns about norms as constraints, without insisting that they be explicitly grounded in a preference-utilitarian welfare analysis. For instance, the Stanford economics department uses a serial dictatorship to allocate office space to graduate students (even though the department is no stranger to the virtues of auctions). An auction would allow graduate students to signal the intensity of their preferences, but could change the egalitarian norms of shared office space. In this case, a concern about egalitarian office norms can be addressed by a 'reduced-form' constraint that prohibits money transfers.

There are also ethical virtues of markets that preference utilitarianism does not capture. Markets give people the freedom to decide what transactions to engage in, what to produce, what to consume, and for whom to work, and this is valuable intrinsically, not just because it satisfies their preferences (Sen, 2001, p. 25-30).

> The discipline of economics has tended to move away from focusing on the value of freedoms to that of utilities, income and wealth. This narrowing of focus leads to an underappreciation of the full role of the market mechanism, even though economics as a profession can hardly be accused of not praising the markets enough. The issue, however, is not the amount of praise, but the reasons for it. (Sen, 2001, p. 27)

Suppose, for the sake of argument, that university students' preferences over course allocations were highly predictable given their demographic characteristics, so that it was not necessary to consult students in order to assign them their (statistically) preferred courses. It would still be valuable to implement a course allocation system that allowed students to make choices and be causally responsible for their course load, even if this did not substantially increase preference satisfaction. Our ethical assessment of a market design depends not just on its consequences, but also on the procedure that leads to those consequences[9]; the correct value function may instead be $v : \mathcal{D} \times \mathcal{C} \to \mathcal{J}$. One virtue of a well-designed market is that it is a free procedure, in the sense that each market participant can freely determine part of the outcome. This virtue could feature more prominently in advocacy for market design.

---

[8]This is not the only time that market design has encountered religion; stable matching mechanisms are used by Hebrew Union College to assign graduating rabbis to congregations (Roth, 2000; Bodin and Panken, 2003). Presumably, not every design for this market is compatible with the social meaning of rabbis.

[9]Equivalently, one can understand consequences in a rich way - a "comprehensive outcome" that takes note of the procedures and choices that got us there (Sen, 1997).

Even if economists are mostly convinced that preference utilitarianism is (approximately) the correct ethical theory, many policymakers are not similarly convinced. Since "[t]he adoption of a design is at least partly a political process" (Roth, 2002), it is valuable for the recommendations of market designers to not depend sensitively on the truth of preference utilitarianism. If a design is desirable *only* on preference utilitarian grounds, then it may be difficult to persuade policymakers that it is desirable. Moreover, even if policymakers were persuaded that only preference satisfaction matters when evaluating designs, they often have to justify their actions to stakeholders with broader ethical concerns. Consequently, market designers should not rely exclusively on preference utilitarianism to evaluate designs.

Indeed, market designers *do not* rely exclusively on preference utilitarianism to evaluate designs. There is a panoply of fairness concepts in the market design literature, such as the equal treatment of equals, envy-freeness, the elimination of justified envy, envy bounded by a single good, maximin shares, and *ex post* fairness (Gale, 1987; Zhou, 1990; Bogomolnaia and Moulin, 2001; Abdulkadiroglu and Sönmez, 2003; Budish, 2011; Budish et al., 2013; Akbarpour and Nikzad, 2015). These concepts constrain the distribution of goods and are not implied by preference utilitarianism, which just requires that we maximize aggregate utility.[10] Papers studying market design for online platforms tend to regard taste-based racial discrimination as regrettable (Edelman et al., 2016; Ge et al., 2016). They do not take the preference utilitarian position that racist preferences are just as worthy of satisfaction as non-racist preferences. Furthermore, market designers recognize the need to consider non-utilitarian ethical requirements (whether or not they believe them), because "distaste for certain kinds of transactions is a real constraint, every bit as real as the constraints imposed by technology or by the requirements of incentives and efficiency" (Roth, 2007).

3 An Expanded Model and a Division of Labor

There is again a set of feasible designs, $\mathcal{D}$, a set of consequences, $\mathcal{C}$, and a function $f : \mathcal{D} \to \mathcal{C}$. Value judgments need not be consequentialist; there is some value function $v : \mathcal{D} \times \mathcal{C} \to \mathcal{J}$.

I have argued that, for the purposes of market design, we should not assume that v is preference utilitarian. One natural course of action is to consider other complete ethical theories so as to find a more firm foundation.[11] We could seek an ethical theory that does for market design what preference utilitarianism does for the rest of economics, a theory that provides a unified normative foundation for policy analysis.

It would be worthy and interesting to seek a complete normative foundation for market design, but it is not an urgent project. Engineering disciplines, from bridge design to chemical engineering to medicine, have made substantial advances without a complete normative foundation. We do not usually pause to ask whether a bridge designer is a preference utilitarian. Indeed, a strong

---

[10]In addition to fairness concepts that depend on the initial and final distributions of goods, one could instead investigate fair transition principles. Thomson (2011) provides a survey.

[11]Nussbaum (2016) observes that "[e]conomists have often built structures of great precision and sophistication, often mathematical sophistication, over foundations that philosophers criticize, and they are loathe to rethink first principles."

methodological allegiance to one ethical theory will make it harder to advise policymakers who do not share that theory. The positive insights of market design will be more believable if they are not bundled with a particular normative doctrine.

Within our expanded model, one can propose the following (rough) division of labor: Market designers should study f. They should invent new designs (learning the domain of f), investigate the graph $(d, f(d))$ which connects designs and consequences, and characterize the achievable consequences (the range $f(\mathcal{D})$). They should not attempt to resolve fundamental ethical questions, leaving the specification of v to policymakers, stakeholders, and ethicists.

There are plausible reasons to recommend this stark division of labor. There is the usual argument from comparative advantage: Market designers are typically adept at mathematical and statistical analysis. Market designers (*qua* economists) have powerful tools to understand cause and consequence in complex markets, but no special expertise in solving ethical conundrums. The problem of choosing an ethical theory can be handled by policymakers (who are often answerable to the relevant stakeholders) or philosophers (who have the benefit of specialized training and long reflection).

Furthermore, it is not clear that any discipline can settle on the correct normative foundation for market design, so it is not reasonable to wait for the question to be resolved before proceeding. It is difficult to conclusively resolve a fundamental ethical question, because they are not amenable to mathematical proof, experimentation, or scientific observation. There are rarely decisive arguments for strong conclusions in philosophy.[12] Consequently, "philosophical arguments typically lead not to agreement but to sophisticated disagreement. Advocates of a view learn what extra commitments they need to take on to avoid the arguments. Bad versions of a view are rejected and sophisticated versions are developed in their place." (Chalmers, 2015)

Moreover, market design can provide normative insights without directly studying v, because some *prima facie* ethical disagreements are really disagreements about f. A layperson might judge that a market in kidneys is exploitative, because she predicts that the market-clearing price will be low and only poor people will sell their kidneys. If it turns out that the clearing price is high, every transaction is bundled with health insurance, and even well-off people sell their kidneys, then she might revise her value judgment; she was mistaken about f. More fundamentally, many ethical judgments depend on what is feasible, because ought implies can (Kant, 1781, p. 548). Therefore, just investigating $f(\mathcal{D})$ can yield normative conclusions. For instance, when designing a school choice system with school-specific priorities, a policymaker might think that it is morally obligatory to implement a system that is Pareto optimal for students and eliminates justified envy. However, this is not possible (Abdulkadiroglu and Sönmez, 2003), so it cannot be a moral obligation. The policymaker is mistaken about $f(\mathcal{D})$, and market design can provide that insight without taking a stand on v. Ethical theories must weigh the trade-offs between different designs, and economic

---

[12]Van Inwagen (2004) writes, "Disagreement in philosophy is pervasive and irresoluble. There is almost no thesis in philosophy about which philosophers agree. If there is any philosophical thesis that all or most philosophers affirm, it is a negative thesis: that formalism is not the right philosophy of mathematics, for example, or that knowledge is not (simply) justified, true belief."

methods can make clear what those trade-offs are.

The foregoing arguments suggest that market design should be neutral between reasonable ethical positions; it should not assume the truth of any one complete ethical theory. However, being neutral between ethical positions is not the same as ignoring them. Market design cannot be completely silent about ethics.

One reason that market design cannot be silent about ethics is that the search space $\mathcal{D}$ is vast. We will never know all there is to know about the graph $(d, f(d))$, except in the simplest of markets. Ultimately, we are interested in designing markets that run well, not markets that run badly - few papers are written about welfare-minimizing market designs. "Well" and "badly" have normative content; they are assessments relative to some goal or standard. Normative criteria guide us about which parts of $\mathcal{D}$ to investigate. If we do not have explicit criteria, we will rely on implicit criteria, often those that are most familiar to us, such as revenue maximization, welfare maximization, or Pareto optimality.

Another reason that market design cannot be silent about ethics is that, ultimately, f and v must be assessed jointly in order to steer policy. We need results about f that can be used to evaluate $v(d, f(d))$. Every consequence is packed with detail, but not every detail matters for evaluating v. Consequently, we must learn the criteria that different ethical theories use to make value judgments: Is the design fair? Does it prevent vulnerable participants from being exploited? Will participants be pleased with the outcome? Does it protect individual liberty? Bridge designers know how to state and optimize for the ethically relevant details of bridges, such as cost, safety, and durability. Consequently, it does not matter whether a bridge designer is a preference utilitarian - it just matters how their proposed design scores on these standard dimensions. The ethically relevant dimensions of markets are substantially richer, and we do not yet know how to state some of them.

Market design must speak about ethics because policymakers sometimes need help expressing what they want. For instance, in a complicated assignment problem, we can formalize what it means to be fair, using criteria such as the Shapley value, equal treatment of equals, or elimination of justified envy (Shapley, 1953; Zhou, 1990; Bogomolnaia and Moulin, 2001; Abdulkadiroglu and Sönmez, 2003). However, these definitions are subtle and might not occur to a layperson. A policymaker may be familiar with the details of their environment, and yet not know how to state their ethical requirements in precise terms. Market designers are familiar with the idea that participants require well-designed reporting languages in order to express their preferences (Milgrom, 2011; Budish and Kessler, 2016); the same observation holds true for policymakers expressing ethical goals and requirements.

4 Informed Neutrality Between Reasonable Ethical Positions

I will expand the model again, and sketch what it means for market design to maintain an informed neutrality between reasonable ethical positions. I do not know whether this proposal is novel, since

there appears to be no prior work that directly addresses the question. It could be that this is already the *de facto* normative framework used by many market designers, in which case this paper will do the small service of turning knowledge into common knowledge.

Market design is unlikely to settle the question, "All things considered, should we adopt this design?" The answer to this will depend on having a theory of normative ethics, and market design should remain neutral between reasonable[13] ethical theories. However, every ethical theory picks out certain features of a situation as morally relevant. It matters whether a design *maximizes welfare*, whether it *is fair*, whether it *is transparent*, whether it *respects individual liberty* and so on. Some features are even picked out as relevant by multiple ethical theories; not all theories hold that what one should do depends *only* on the welfare consequences, but most plausible theories agree that welfare matters. Market designers should formalize features such as these and study them.

There is a set of designs $\mathcal{D}$, a set of consequences $\mathcal{C}$, a set of *intermediate* judgments $\mathcal{I}$, and a set of *overall* normative judgments $\mathcal{J}$. Each design results in some consequence, $f : \mathcal{D} \to \mathcal{C}$. Each design-consequence pair results in some intermediate judgment, $g : \mathcal{D} \times \mathcal{C} \to \mathcal{I}$. Each intermediate judgment implies some overall judgment, $v : \mathcal{I} \to \mathcal{J}$.

An intermediate judgment specifies normatively relevant features of a design-consequence pair. Examples of intermediate judgments include:

1. "d produces the doctor-optimal stable matching, which is also the hospital-pessimal stable matching."

2. "d maximizes welfare and respects individual liberty, but does not preserve privacy."

3. "d is approximately efficient and *ex ante* fair, but not *ex post* fair."

4. "d gives every donor in a kidney donation chain the freedom to renege on their promise, though most donors decide to continue the chain."[14]

An overall judgment takes the form, "all things considered, we should implement d", or "all things considered, we should not implement d".

Informed neutrality means the following division of labor: Market designers should study f, investigating its domain, range, and graph. They need not seek to resolve fundamental ethical questions. Instead, they should understand reasonable ethical theories; that is, plausible specifications of v, at least enough to know what intermediate judgments they take as inputs, and how to formalize these concepts without doing too much violence to their ordinary-language meaning.[15] With

---

[13]It is difficult to define a "reasonable ethical theory". Roughly, I mean something in the spirit of what Rawls (1993) calls a "reasonable comprehensive doctrine": (1) It "organizes and characterizes recognized values so that they are compatible with one another and express an intelligible view of the world", (2) it singles out "which values to count as especially significant and how to balance them when they conflict", and (3) it evolves in light of "good and sufficient reasons".

[14]For more on kidney donation chains, see Roth et al. (2006).

[15]To this end, some degree of training in ethics may be necessary, at least for the market designers doing the formalizing. Consider the example of medicine: Some doctors think especially about ethics and formulate common standards. Others rely on these standards in the course of their work.

that in mind, they should formalize g and $\mathcal{I}$, and seek results about f that inform the assessment of $g(d, f(d))$.

Via this approach, market design can offer normative guidance without prescribing an entire ethical theory. To a policymaker concerned about designing a fair market, we could ask, "What kind of fairness do you mean?", and offer a range of plausible definitions, along with results that relate fairness to other properties the policymaker might care about, such as efficiency and incentives. (The extensive literature on fairness constraints suggests that informed neutrality is at least a feasible way to design markets.) Instead of constraining market design to focus only on positive results, informed neutrality prescribes a different division of labor: In addition to studying cause and effect in markets, economists also have a comparative advantage in *stating precisely the normatively-relevant properties of complex systems*, as is exhibited by the foundational work on social choice (Arrow, 1963; Sen, 1970a).

There are other normatively relevant properties, in addition to welfare and fairness, that might be usefully formalized when studying market design. (The list that follows is not exhaustive; it points to open questions for future research.) Some ethical theories require us to assess whether a design *protects vulnerable agents from exploitation* (Satz, 2010); this seems to restrict what transactions can occur when an agent's outside options are sufficiently bad, or require the design to insure agents against bad outside options. *Individual privacy* can be valuable for intrinsic as well as instrumental reasons, and so is worth investigating (Izmalkov et al., 2005; Milgrom and Segal, 2017). Market mechanisms can often be complex; they must sometimes be made *simple* and *transparent* in order for people to participate as moral equals (Barthe et al., 2016; Li, 2016, 2017). Participants should be able to answer questions such as, "What should I do to get what I want?" and "Why did I get what was allocated to me?" Finally, it matters whether a design *respects individual liberty*; Sen (1970b) offers a definition of liberalism in social choice, but other definitions may be appropriate for market design.

A complete normative foundation specifies a single ethical theory, on the basis that the theory is true. By contrast, informed neutrality admits plural ethical criteria, on the basis that those criteria are plausible. Whether an ethical criterion is plausible depends on whether it is coherent, survives sustained reflection, and accords roughly with our intuitions. It also depends on whether policymakers and relevant stakeholders care about that criterion, which is an empirical question that can be addressed by observation and experiments (Ambuehl et al., 2015; Ambuehl and Ockenfels, 2017; Sytsma and Buckwalter, 2016).

5   A Trialogue

The following fiction illustrates how informed neutrality might play out in practice. It is a conversation between three characters: Solomon, a policymaker, seeks advice from two market designers, Jeremy and Isaiah. Jeremy is a preference utilitarian of the textbook variety.[16] Isaiah practices

---

[16]Though, of course, his namesake was a utilitarian of a different stripe.

informed neutrality between reasonable ethical positions.

Solomon: I want to design an auction for widgets. There's a seller with one widget, and two buyers. I care about the seller's revenue and about efficiency. What should I do?

Jeremy: I'm reluctant to trade off revenue and efficiency, since it's difficult to make inter-personal comparisons of utility. Still, assuming that a dollar gives the same marginal utility to every participant, any efficient auction maximizes utility. If you care about revenue in addition to efficiency, it must be that you regard a dollar to the seller as being worth more than a dollar to a buyer. In that case, we should figure out the correct exchange rate before proceeding.

Solomon: This is all very strange. I hadn't even thought about maximizing utility, and as to figuring out an exchange rate - I wouldn't know where to begin.

Isaiah: It needn't all be about utility. There are other reasons to care about revenue - for instance, the seller has worked hard to produce the widget, and may deserve compensation for his labors. However, we needn't settle that argument today. Under some standard assumptions[17], we face the following trade-off: A fully efficient auction requires that the object always go to the buyer with the highest value. But we can raise revenue by implementing a reserve price. Up to a certain point, the higher the reserve price, the higher is revenue, but the lower is efficiency.

Solomon: I see. So I can increase revenue by having a reserve price, but this means that the widget sometimes won't be sold, even though it would be efficient to sell it. Still, I'm wary of those "standard assumptions". I think that bidder A is much more likely to have a high value than bidder B; how does that change your recommendations?

Jeremy: Well, assuming that you care even a bit about revenue, you should treat the bidders asymmetrically so as to increase competition.[18] For instance, if you are running a first-price auction, you could use a scoring rule that allows B to win even when his bid is slightly lower than A's bid.

Solomon: That seems rather unfair -

Jeremy: Fairness doesn't come into it. Assuming we can compare utilities across persons, we should maximize some (weighted) sum of utilities. Do you really mean that we have the weights wrong?

Isaiah: I think Solomon means something else. There are many ways that an auction can be fair or unfair. On the one hand, your scoring rule violates the equal treatment of equals - A has to bid $110 to beat B's bid of $100, but not *vice versa*. It is also not envy-free - A would rather occupy B's position than his own. On the other hand, your scoring rule makes the (*ex ante* expected) utilities of the bidders more equal, so in that sense it is fairer than a symmetric rule. Solomon, you've always been in favor of equal division. Which kinds of fairness matter to you?

Solomon: I forgot to mention - bidder A wants the widget more because he needs it for a medical device that will cure his chronic illness. I'll have to think more about fairness, but I have one last

---

[17]Isaiah is thinking of the symmetric independent private values model (Myerson, 1981).

[18]Jeremy is assuming that the distributions satisfy $F_A(v + k) = F_B(v)$ for some constant k, and that the hazard rates are increasing, which implies that the virtual value of agent A with type v is lower than the virtual value of agent B with type v.

question. Both bidders tell me that they're concerned about privacy - they would rather keep secret their values for widgets. How should this affect our design?

Jeremy: Well, first we need to figure out the bidders' utility-from-privacy, in order to maximize utility. I can design a clever mechanism to elicit this before the auction - if the bidders truly desire privacy, they should be willing to pay for it.

Isaiah: And what if the bidders want privacy in that mechanism too? The regress is infinite. Solomon, if you think that it's reasonable for the bidders to demand privacy, perhaps we should take privacy as a constraint without cashing it out in terms of utility. Now, obviously we can't guarantee total privacy for every bidder. In any auction, bidders will make choices, and those choices will be informative about their values. But we can preserve the privacy of the winning bidder, by running a kind of ascending clock auction.[19] However, preserving winner privacy prevents us from running asynchronous procedures, and thus raises logistical costs.

6   Conclusion

In this paper, I addressed the question, "How, if at all, should economic engineers think about ethics?" I argued for three theses: Firstly, that the literature on market design does not, and should not, rely exclusively on preference utilitarianism to evaluate designs. Secondly, that market designers should study the connection between designs and consequences, and should not attempt to resolve fundamental ethical questions. Thirdly, that the theory and practice of market design should maintain an informed neutrality between reasonable ethical positions.

Informed neutrality requires us to engage with reasonable ethical theories, not to resolve which one is correct, but to formalize the features that they regard as morally relevant. Market design already admits non-utilitarian ethical criteria, and we should expand this practice. After all, we need to be precise about the ethical trade-offs between different market designs, and "any argument which is carried out with sufficient precision is mathematical" (Gale and Shapley, 1962).

Informed neutrality is a framework for market designers to engage with ethics. However, it could be that the preceding arguments have a wider scope: Why should market designers engage with ethical *theories* in particular, rather than engage with ethical concerns in other ways? For instance, market designers could talk to stakeholders to understand their ethical perspectives, or examine case studies that describe ethical concerns with existing markets. This is a fruitful approach and worth pursuing. For instance, Titmuss (1970) studies market and non-market systems for blood transfusions (in the USA and the UK), and Arrow (1972) offers a critique. However, talking to stakeholders and reading case studies are not mutually exclusive with engaging with ethical theories. One advantage of theory is that it creates a general conceptual framework that applies to many cases. Concepts such as stability or incentive compatibility are powerful precisely because they are abstract. So too with ethical concepts: We need not reinvent the wheel for each new market, so it

---

[19]In fact, this is the only way to do so (Milgrom and Segal, 2017).

is desirable to understand ethical theories; that is, systems for ethical judgment that are not bound to a narrow context.

Equally, why should informed neutrality be a normative framework for market design in particular, rather than for economics in general? This stronger thesis could be true, but it is beyond the scope of this paper to argue for it. Notably, economics typically considers self-interested agents trading their private property. As the perennial quotation goes, "It is not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest." (Smith, 1776) By contrast, modern market design often facilitates transactions in situations that were not previously thought of as transactional. Market designers have proposed using the Top Trading Cycles algorithm to allocate students to public schools (Abdulkadiroglu and Sönmez, 2003), even though priority at a school is not (legally speaking) the private property of each student.[20] Market designers have created clearinghouses for kidney exchange (Roth et al., 2005), and the donors in these exchanges participate (at least partly) out of benevolence. It is possible that that preference utilitarianism is approximately the right ethical theory for 'standard' markets, but does not quite fit the new contexts that market designers now encounter. Alternatively, it could be that preference utilitarianism is a good ethical theory for these novel contexts, even though economists were not thinking about these contexts when we adopted the theory. That would, at least, be surprising.

References

Abdulkadiroglu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *The American Economic Review*, 93(3):729–747.

Akbarpour, M. and Nikzad, A. (2015). Approximate random allocation mechanisms. *working paper*.

Ambuehl, S., Niederle, M., and Roth, A. E. (2015). More money, more problems? can high pay be coercive and repugnant? *American Economic Review: Papers and Proceedings*, 105(5):357–360.

Ambuehl, S. and Ockenfels, A. (2017). The ethics of incentivizing the uninformed. a vignette study. *American Economic Review: Papers and Proceedings, forthcoming*.

Arrow, K. J. (1963). *Social Choice and Individual Values*. Yale University Press.

Arrow, K. J. (1972). Gifts and exchanges. *Philosophy & Public Affairs*, pages 343–362.

Barthe, G., Gaboardi, M., Arias, E. J. G., Hsu, J., Roth, A., and Strub, P.-Y. (2016). Computer-aided verification for mechanism design. In *International Conference on Web and Internet Economics*, pages 279–293. Springer.

Bodin, L. and Panken, A. (2003). High tech for a higher authority: The placement of graduating rabbis from hebrew union college: Jewish institute of religion. *Interfaces*, 33(3):1–11.

---

[20]One could not, for instance, sign a contract selling one's priority for money.

Bogomolnaia, A. and Moulin, H. (2001). A new solution to the random assignment problem. *Journal of Economic theory*, 100(2):295–328.

Bourget, D. and Chalmers, D. J. (2014). What do philosophers believe? *Philosophical Studies*, 170(3):465–500.

Budish, E. (2011). The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*, 119(6):1061–1103.

Budish, E., Che, Y.-K., Kojima, F., and Milgrom, P. (2013). Designing random allocation mechanisms: Theory and applications. *American Economic Review*, 103(2):585–623.

Budish, E., Cramton, P., and Shim, J. (2015). The high-frequency trading arms race: Frequent batch auctions as a market design response. *The Quarterly Journal of Economics*, 130(4):1547–1621.

Budish, E. and Kessler, J. (2016). Bringing real market participants' real preferences into the lab: An experiment that changed the course allocation mechanism at wharton. *working paper*.

Catholic Church (2000). *Catechism of the Catholic Church.* Vatican: Libreria Editrice Vaticana, second edition.

Chalmers, D. J. (2015). Why isn't there more progress in philosophy? *Philosophy*, 90(01):3–31.

Cross, F. L. and Livingstone, E. A. (2005). *The Oxford Dictionary of the Christian church.* Oxford University Press, USA.

Edelman, B., Luca, M., and Svirsky, D. (2016). Racial discrimination in the sharing economy: Evidence from a field experiment. *forthcoming, American Economic Journal: Applied Economics*.

Gale, D. (1987). College course assignments and optimal lotteries. Mimeo, University of California at Berkeley.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.

Ge, Y., Knittel, C. R., MacKenzie, D., and Zoepf, S. (2016). Racial and gender discrimination in transportation network companies. Technical report, National Bureau of Economic Research.

Izmalkov, S., Micali, S., and Lepinski, M. (2005). Rational secure computation and ideal mechanism design. In *Foundations of Computer Science, 2005. FOCS 2005. 46th Annual IEEE Symposium on*, pages 585–594. IEEE.

Kant, I. (1781). *Kritik der reinen Vernunft.* Johann Friedrich Harknosh, Riga.

Krishna, V. (2010). *Auction Theory.* Academic Press.

Lee, S. and Niederle, M. (2015). Propose with a rose? signaling in internet dating markets. *Experimental Economics*, 18(4):731–755.

Li, S. (2016). Obviously strategy-proof mechanisms. *working paper*.

Li, S. (2017). Obvious *Ex Post* equilibrium. *American Economic Review: Papers and Proceedings, forthcoming*.

Milgrom, P. (2011). Critical issues in the practice of market design. *Economic Inquiry*, 49(2):311–320.

Milgrom, P. and Segal, I. (2017). Deferred acceptance auctions and radio spectrum reallocation. *working paper*.

Myerson, R. B. (1981). Optimal auction design. *Mathematics of operations research*, 6(1):58–73.

Nussbaum, M. C. (2016). Economics still needs philosophy. *Review of Social Economy*, 74(3):229–247.

Parfit, D. (1984). *Reasons and Persons*. Oxford University Press, Oxford.

Rawls, J. (1993). *Political Liberalism*. John Dewey essays in philosophy. Columbia University Press.

Roth, A. E. (2000). *Game Practice: Contributions from Applied Game Theory*, chapter Game Theory as a Tool for Market Design., pages 7–18. Kluwer Academic Publishers.

Roth, A. E. (2002). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica*, 70(4):1341–1378.

Roth, A. E. (2007). Repugnance as a constraint on markets. *The Journal of Economic Perspectives*, 21(3):37–58.

Roth, A. E. and Peranson, E. (1999). The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review*, 89(4):748.

Roth, A. E., Sönmez, T., and Ünver, M. U. (2005). A kidney exchange clearinghouse in New England. *American Economic Review*, pages 376–380.

Roth, A. E., Sönmez, T., Ünver, M. U., Delmonico, F. L., and Saidman, S. L. (2006). Utilizing list exchange and nondirected donation through chainpaired kidney donations. *American Journal of transplantation*, 6(11):2694–2705.

Sandel, M. J. (2012). *What Money Can't Buy: The moral limits of markets*. Macmillan.

Satz, D. (2010). *Why Some Things Should Not Be For Sale: The moral limits of markets*. Oxford University Press.

Sen, A. (1970a). *Collective Choice and Social Welfare*. North-Holland, Amsterdam.

Sen, A. (1970b). The impossibility of a paretian liberal. *Journal of Political Economy*, 78(1):152–157.

Sen, A. (1987). *On Ethics and Economics*. Blackwell, Oxford.

Sen, A. (1997). Maximization and the act of choice. *Econometrica*, pages 745–779.

Sen, A. (2001). *Development as Freedom*. Oxford University Press.

Shapley, L. S. (1953). A value for n-person games. In Kuhn, H. and Tucker, A., editors, *Contributions to the Theory of Games*, volume 28 of *Annals of Mathematical Studies*, pages 307–317. Princeton University Press, Princeton, New Jersey.

Smith, A. (1776). *An Inquiry Into the Nature and Causes of the Wealth of Nations*. Number v. 1 in An Inquiry Into the Nature and Causes of the Wealth of Nations. Strahan.

Sytsma, J. and Buckwalter, W. (2016). *A Companion to Experimental Philosophy*. John Wiley & Sons.

Thomson, W. (2011). Fair allocation rules. In Arrow, K. J., Sen, A., and Suzumura, K., editors, *Handbook of Social Choice and Welfare*, volume 2, chapter 21, pages 393–506. Elsevier, Oxford, UK.

Titmuss, R. (1970). *The Gift Relationship: From Human Blood to Social Policy*. Reprinted by the New Press.

Van Inwagen, P. (2004). Freedom to break the laws. *Midwest Studies in Philosophy*, 28(1):334–350.

Von Neumann, J. and Morgenstern, O. (1953). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, New Jersey.

Wengert, T. J. (2015). *Martin Luther's Ninety-Five Theses: With Introduction, Commentary, and Study Guide*. Fortress - via Project MUSE, Minneapolis, MN.

Zhou, L. (1990). On a conjecture by gale about one-sided matching problems. *Journal of Economic Theory*, 52(1):123–135.