An overview of practical exchange design

R. Preston McAfee* and Sergei Vassilvitskii

Google, 1600 Amphitheatre Parkway Building 43, Mountain View, CA 94043, USA

We consider the problem of designing an online exchange. We identify the goals of exchange design, and present key techniques for accomplishing these goals along with the tradeoffs inherent in the choices.

Keywords: Exchange design, goals and key techniques, online exchange, selling mechanisms.

THERE is a large literature on the design of selling mechanisms, which has led to a practical methodology for guiding sellers in the choice of selling methods (see note 1). This methodology is related to the theory of mechanism design in the same way that engineering is related to physics: the practical methodology includes not just theoretical insights, but also the results of experimental and practical tests and a growing body of evidence from deployment, including the sale of over $100 billion in radio spectrum, primarily for cellular telephony, as well as in medical residencies, school assignment, electricity, pollution permits and natural resources. While the literature on practical selling and matching is rich indeed, little has been written on the methodology of exchange design, a hole this article begins to fill.

Exchange design differs from selling method design in a number of ways. First, in the design of selling methods, competition between sellers is generally ignored. Ignoring seller competition is reasonable when selling radio spectrum, which is offered by a government monopoly, but unreasonable in many other settings, such as eBay auctions. Selling methods typically favour the seller, while exchange design should balance the needs of both sides of the market. Second, there are few articles on the theory of exchange design; a methodology must therefore piece together the principles from one-sided design theory and rely more heavily on insights developed either through selling method practice or through experimental insights. Third, some selling methods simply do not apply: an ascending auction is not a sensible design in a two-sided market. Fourth, some issues relevant to an exchange environment, such as whom to charge, are moot in the one-sided auction environment.

To focus our ideas, we consider exchanges in which a group of sellers sell a sequence of differentiated goods to a group of buyers. A typical eBay exchange of, say, a specific model of Apple iPod would serve as an illustrative example. In this case, some sellers appear once to sell one iPod and others run businesses and repeatedly come to the market with new iPods. Sellers differ in the value they place on the items and potentially on how they discount the future. Some items, like tickets for seats in baseball games or advertising impressions on web pages, expire at a fixed time, whereas others, like baseball cards, remain valuable indefinitely, but may be costly to store. Many buyers purchase a single unit, but some buy many units or be repeat buyers. Buyers may need the item by a specific time – a camera for a vacation – or may be flexible.

Broadly speaking, an exchange is a mapping from expressed preferences of participants into an allocation of goods and money. Thus, buyers and sellers will express their preferences or valuations, through a process that may be iterative and reactive, and a rule will determine who gets what and who pays whom how much. The process and methods of expressing preferences and the mapping into allocations is exchange design.

Goals of exchange design

We identify four goals of exchange design. A well-designed exchange should be: efficient, expressively easy, strategically simple and neutral.

Efficiency

Efficiency entails maximizing the total value generated by the exchange, that is, maximizing the gains from trade. Efficiency is valuable not only because it is socially good, providing the greatest total benefit, but also because it wards off competing exchanges. An exchange that operates efficiently should out-compete exchanges that do not because the profits for participation are overall higher under efficiency (see note 2).

Exchanges are costly to operate; another aspect of efficiency is *operational efficiency*, which involves minimizing the costs of operating the exchange. Many of the choices that potentially improve the efficiency of transactions increase the cost of operation, and thus there is a conflict between transactional and operational efficiency.

Expressiveness

An important tradeoff in designing exchanges is the language participants use to express their preferences. In

---

*For correspondence. (e-mail: preston@mcafee.cc) choosing a transaction language, it is valuable to let participants say what matters to them. Thus, if the colour of an iPod matters, sellers need to be able to post the colour and buyers to condition on it; otherwise buyers will fail to obtain the item they want and hence be less willing to pay. Efficiency suffers when the buyer preferring a blue iPod gets a green one and vice versa.

At the same time, greater expressiveness comes with two distinct costs. First, all parties face greater complexity, involving both a time cost in understanding the language and a greater probability of mistakes. Second, markets become thinner, tending to increased price variance and reducing both buyer and seller information within a category.

A key insight, due as far as we know to Paul Milgrom, is that participants need not use the same language (see note 3). Thus, some buyers might go to bid on an iPod, and need not concern themselves with colour, if in fact colour does not matter to them. Others may choose to bid only on a blue iPod. The system then figures out which buyer gets which iPod based on efficiency. From a programming perspective, distinct participants may use different interfaces or 'front ends' in which to express their preferences. The exchange may not only offer different interfaces for buyers and sellers, but different interfaces for distinct buyer types. Buyers may select the interface that best suits their needs; behind the scenes ('back end') processing then aggregates the preferences into a unified language. The beauty of this insight is that an expressively complex language may have an expressively easy version for participants who do not need the full power of the language. Using multiple interfaces is especially important when new users need simplicity, while experienced users demand more control.

A language may be adequate even when it does not cover all the things that matter, provided the uncovered distinctions do not matter too much. For example, most airlines do not price the middle seat separately from aisle or window seats, even though the middle seat is less desirable. Note that even if the value difference is 5%, the actual loss in efficiency is likely much less, a topic addressed below under price deltas.

Strategic considerations

If the optimal selection of choices in one game requires less information or guesses than in another game, the first game is strategically simpler. The classic example is the Vickrey auction, in which the high bidder pays not his own bid, but the second highest bid. This game is strategically simpler than the first price (pay your bid) auction, although the reason is widely misunderstood. It is generally not optimal to bid one's estimate of value; bidding one's value is only optimal in the case when one knows one's value. In the more typical case, other bids are a signal of value. In such a setting, the second price auction remains strategically simpler than a first price auction, because it does not require forecasting the second highest bid, just asking what one would be willing to pay if one tied with the second highest bid.

Strategic simplicity is valuable for two distinct reasons. First, simplifying the problem faced by participants lowers their cost of participating, thus encouraging participation. Second, strategic simplicity reduces the chance of regret and mistakes, improving the efficiency of the exchange. Both of these reasons imply that strategic simplicity will encourage more aggressive participation. However, there is often a tension between strategic simplicity and expressiveness. A game which randomly assigns goods to buyers based on their willingness to pay a fixed price is strategically simple – buyers just say yes or no – but not adequately expressive to achieve efficiency.

Neutrality

Exchanges may be designed to favour a specific party. The *National Residency Matching Program*[11], which places physicians in hospital residencies, was slightly tilted in favour of the hospitals. Enron's exchange *EnronOnline* was tilted in favour of Enron in that buyers and sellers could only trade with each other indirectly; all trades were with Enron. The presence of a bias is usually costly to the exchange in terms of efficiency; as a concept, efficiency is unbiased. A biased exchange is therefore vulnerable to an unbiased exchange, because at least some participants will prefer the unbiased exchange. As we use the term, fairness, neutrality and unbiasedness are all synonyms; the exchange should put efficiency first and not favour specific parties.

In some cases, it may be necessary to encourage specific parties who would not otherwise participate. For example, large electricity users have better alternative means of obtaining electricity through self-generation than smaller users and will join a marketplace only under relatively advantageous terms. In this case, it is desirable to provide equal access to favourable treatment, but set prices so that only the large participants find it in their interest to choose the 'favourable' treatment. For example, a large fixed fee with a low marginal price is attractive only to large buyers.

The value of neutrality extends beyond market efficiency. Participants that consider themselves poorly served will attempt to engineer better deals. For example, small participants might join a consortium and try to access better treatment. Such efforts are better deterred by fair treatment, since the efforts themselves are costly.

The use of data in exchange design

There is a vast amount of information used by the participants in an exchange. This includes data supplied by buyers and sellers, data provided by the exchange itself, and data brought by third parties and used by buyers or sellers. How an exchange handles the data of these parties is critical to efficiency and other goals.

Revealing information and transparency

Revealing information is an integral part of exchange design. There are various categories of information that are helpful both to market efficiency and speed of convergence. First, for the same reasons that it is helpful to provide information about product quality in one-sided auctions (it prevents adverse inferences and makes beliefs more symmetric and hence markets more competitive), it is advantageous in two-sided markets as well[12]. In large anonymous markets where the identity of the counterparty is not known, reputation scores, like eBay's and Amazon's user ratings, or hotel ratings visible on Priceline increase efficiency by allowing the participants to adjust their bids in response.

Second, it is useful to publish marketplace statistics. Such statistics, like the Dow Jones Industrial Average or the West Texas Intermediate Crude oil price, serve an important informational role even for market participants not directly involved in the trade of those items. Marketplace statistics provide summary information on the state of an exchange. For example, if I am trading baseball cards on a collectibles exchange, I might want to know how antique presidential buttons are doing, not because antique presidential buttons are directly relevant, but because some of the same forces that influence the prices of baseball cards affect presidential buttons as well.

It is important that marketplace statistics have the right level of aggregation. Make them too aggregate – GDP of the world – and the statistics will not be helpful to traders. Make them too detailed – price of a Barry Bonds baseball card – and the data on which the statistic is based will have a lot of random noise and, worse still, be subject to manipulation by individual traders. Consequently, the target is to make the statistics broad enough to cover thickly traded markets that matter, and no broader. If the first thing participants do with the statistics is aggregate them into useful summaries, the statistics were too finely divided.

Marketplace statistics and data about the specific opportunity being traded are usefully provided by the exchange itself. For example, the US Department of the Interior provides information about the oil tracts that it sells. This provision of seismic and other information is efficient because otherwise the cost of collecting the information would be duplicated by many or all participants. Moreover, providing information tends to increase the average price of the sales[12].

Revealing information about specific participants generally discourages participation and is often unwise. However, published quality scores, such as user rankings, are an exception as they can facilitate exchange.

Both transparency and marketplace statistics are enhanced using uniform price auctions, meaning all parties buying or selling the same thing at the same time have the same price. Uniform pricing makes it easier to publish meaningful price statistics, and reduces the gaming that might otherwise arise in pursuit of differentially attractive prices.

Iteration and adaptive bidding

Iterated mechanisms – where participants are able to state a simple form of preferences (e.g. bids) and then revise them in light of either the tentative allocation or others' statements – generally simplify the strategies as well as improve expressiveness. Iterated mechanisms are distinguished by whether past statements are binding or are treated like offers which are supplanted by alternate offers.

One example of why iterated mechanisms are desirable can be found in the periodic sale of offshore oil leases in the Gulf of Mexico by the U.S. Department of the Interior. Generally one to two hundred leases are offered simultaneously, by a simultaneous sealed-bid auction. Consider a bidder who would like to win ten leases, preferably near each other. This goal is essentially impossible to implement, for it would require forecasting what the other bidders will bid on. Moreover, the desire to minimize payment also requires forecasting what the other bidders will bid. As a consequence, bidders in this auction face a great deal of downside risk. If bidders diversify and bid on more than their target purchases, they risk buying more than desired. If they bid on only what they desire, they risk getting too few. The result is a mess; half the items get no bids at all and only half the sold units are ever developed[13]. An iterated version of the auction, with binding bids (buyer bids cannot be lowered) would be much more effective at generating high prices and efficient allocations.

Generally binding statements are useful because commitment gives force to bids and statements of preference. Binding will slow the mechanism – participants will hesitate to make aggressive offers that cannot be withdrawn – but binding statements permit the mechanism to use 'successive improvement', where each step is welfare-improving because it optimizes not only on the current bids, but over all past bids as well.

In some cases both binding and non-binding statements are useful. For example, binding statements in the form of price offers can be combined with a scratchpad, wherein participants attempt to cooperate. Thus, if a buyer needs two items and is unwilling to bid for the items separately (e.g. if they have little value by themselves), a seller of one item might want to ask other sellers to propose prices for the second item, coordinating their offers to satisfy the buyer. Such a scratchpad is a useful communication tool in many instances, but like any communication tool available to one side of a market, a scratchpad is an invitation to collusion, even when the language is tightly controlled. For example, participants in the US FCC auctions allegedly used trailing digits of the bids themselves to communicate; as the companies were all telephone companies, the participants were familiar with telephone keypads and would naturally translate 288 to ATT (ref. 14). While the allegation of collusive communication was never proved, the possibility of communicating even in a tightly controlled language is clearly demonstrated.

Machine learning

Modern advertising exchanges, like Google's Adsense and Yahoo!'s RMX, auction off the opportunity to show an advertisement to a user in real time. Since every user visit is unique in some way, determining a precise value for each impression is a daunting task. Even characterizing each advertising opportunity by the user's gender, time of day, webpage content, etc. leads to many trillions of distinct items. To battle this heterogeneity of the users, bidders have turned to machine learning to forecast the value of each impression (based on, for example, the user's propensity to click on the advertisement).

Machine learning is a data driven-field, and often simple algorithms with lots of data outperform more complex procedures that only have a limited amount of training data[15]. The party running the exchange has access to more data than any of the participants and thus is in the best position to provide machine-learned estimates or statistics to all of the participants. In the case of advertising exchanges, Yahoo!'s RMX provided a clickthrough rate prediction service, estimating the probability that the specific user would click on each advertisement.

Unlike the exchange itself, each individual participant does not usually have access to all of the transactions that take place on the exchange, and is thus likely to engage in online learning, continually adjusting the bid based on past performance. The bidder's actions may look irrational as he continually trades off between exploring – finding new successful bidding strategies – and exploiting – realizing the gains from previously successful approaches. Any additional information that the exchange can provide, whether by publishing statistics or engaging with a third party to provide additional classification categories, serves to decrease the efficiency loss due to continued exploration.

Implementation

The four goals of exchange design are often in conflict with each other, and any implementation of an exchange leads to a tradeoff between the competing objectives. Below we describe some of the common tradeoffs and give a historical perspective on how different markets addressed them.

Simple versus complex mechanisms

A major tension in the design of exchanges is transparency. A good example of this tension is the Vickrey–Clarke–Groves mechanism, or VCG (see e.g. ref. 4). VCG works by charging each participant the social cost of their participation. Consider an estate auction. The estate has a variety of things for sale – furniture, two televisions, silverware, a dining room table and chairs, a car and so on. In one operation of the VCG mechanism, each bidder would report his values for all the different sets of items. Some of the items are related in value, such as the table and chairs, and each participant might provide the price of not just the individual items but also groupings. Still other items might be substitutes, so that a participant is willing to pay so much for either television, but does not value a second television nearly as much as the first. All of this information is submitted to the VCG mechanism, which returns both an allocation – perhaps you get one television and the table and chairs – and a price. The price is independent of your reported values, depending instead only on what the items would have fetched in your absence, a 'second price'. The same mechanism works with both buyers and sellers present.

The VCG mechanism has some wonderful properties – it is efficient and it provides a strong incentive for honest reporting of values. However, it is rarely used in practice, at least in part because it is not straightforward to understand (see note 4). While the best course of action for a participant in VCG is in fact to report values honestly, this is hardly obvious to casual observers. Participants generally try to game VCG in experiments, and while this fails to be profitable, it also creates an efficiency loss.

In many situations, there is a tradeoff between transparency and complexity. Complex exchanges may be desirable because they permit sophisticated or subtle trades, but they come at a substantial cost. Participants find optimization difficult or impossible, because the rules of the game are obscure. When facing a black box, participants will probe and test and adjust, often at low participation levels at least initially. Moreover, rumours of maltreatment, bias and cheating will persist since there is little way to combat them.

Conceptual complexity – hard to understand mechanisms – should always be viewed as bad, with the understanding that solutions to more complex problems are costlier to achieve and are more prone to errors. In particular, complexity that serves no efficiency role should be avoided. For example, the US Federal Communications Commission (FCC) adopted a bizarre 'accelerator rule' to adjust the size of the bid increment in its spectrum auctions; the rule was sufficiently complex that the US FCC itself was unable to provide the algorithm to participants in a timely fashion. This rule solved a non-problem, of letting the price deltas (discussed below) vary depending on how much competition was present, at a substantial loss in predictability of the price paths, as well as breaking extant bidding tools. The right policy is to permit increased complexity only when there is a clear efficiency gain.

In some cases, complex mechanisms are necessary because the problem is intrinsically complex. For example, loading the space shuttle with several experiments is an inherently complex problem because adding an experiment may displace a much larger experiment; the complexity is generated by the packing problem. Similarly, the problem of operating trains on a given set of tracks is inherently complex: tracks may be used for either direction, though not at the same time, and slow trains may need to pull off to accommodate faster trains[17]. In these cases, the nature of the allocation problem forces complexity, and hence there is no way for a fully transparent exchange to be used.

When complexity is unavoidable, aggregate statistics can be helpful to assure participants that they are doing well. For example, the pricing of tracks in the train allocation problem might be reported by track segment and by time of day; a participant paying more than the average then has the option of moving to an alternate time or accepting that he needs a particularly expensive time. Similarly, the price per cubic foot on the shuttle might be reported; the owner of the experiment for which a high price was demanded by the system might try to repackage or reshape to see if a lower price can be achieved, whereas the owner of a low-priced experiment need not waste effort probing the system. Aggregate statistics may mitigate non-transparency problems of complex market designs.

Price deltas

A price delta is the smallest permitted price change; in the context of an auction, a price delta is known as a bid increment. Prior to 1997, the smallest change on the New York Stock Exchange was an eighth of a dollar. This was changed to a sixteenth in 1997 and eventually to a penny. A great deal of discussion swirled around this decision; similarly, there is great resistance by consumer groups against the elimination of the US penny. However, economically, small price deltas are not necessary to achieve efficiency. A price delta only matters when the difference in value between the holder or owner of an object and the potential buyer or recipient of an object is less than the price delta. Thus, for small price deltas, the probability that the price delta matters will be proportional to the price delta, and the amount of potential damage is no greater than the price delta. Consequently, the efficiency cost of the price delta is about the price delta squared. For a 10% price delta, the damage is on the order of 1%, generally not a large amount. In contrast, large price deltas imply fast convergence of iterated mechanisms, both because step sizes are larger and because the target is larger.

Mushing

Analogous to price deltas, it is sometimes desirable to treat similar but distinct objects as identical, if this simplifies the problem faced by participants. In the spectrum auctions in Mexico, which were an iterative auction analogous to the US FCC auctions, bidders could substitute between similarly sized bundles of licenses. If the requirement of 'similar sizes' were to be enforced exactly, no substitution would be possible. The solution was to 'mush' the notion of similarity. Specifically, the strategy was to set Mexico City as a size 24 and then make other property sizes relatively simple to add up to 24. All other properties were given sizes 1, 2, 3, 4, 6 or 12. That way, there were many ways to reach 24. If, on the other hand, one of the properties had a size 17, or Mexico City were given a size of 37, the balancing of groups would be much more difficult. In contrast, the US FCC insisted on using sizes based on population to the nearest thousand, making substitutions nearly impossible.

As a strategy, mushing involves the tradeoff between treating unlike objects as alike, but simplifying both the language and the action set facing participants. Some mushing is almost always desirable, while complete mushing – treating diamonds and water chestnuts as the same object – is not. Klemperer[18] presents an alternative strategy for handling the complexity of many items, which will be useful when the product differentiation is substantial and the number of items not prohibitively large, as in the financial circumstances considered.

Assembling buyers

With durable goods, there is a question of whether to sell now or wait for additional buyers to appear. The general solution to this problem is to set a reserve or minimum price on the good, and sell provided the buyer is willing to pay the reserve. With differentiated products, the reserves could vary across the goods, or be common and either give a random assignment or give the buyer a choice of which existing goods to purchase. A sensible and efficient reserve is the expected value of a future buyer discounted to the present.

Efficiency across time requires comparing \$1 today to another amount in the future. The normal method of doing so is to calculate the present value of future dollars by discounting by interest rates. Thus, at an interest rate of 10%, \$110 in one year is equal to \$100 today and vice versa. Then an efficient system maximizes the expected present value of trades.

With undifferentiated, storable goods, efficiency usually requires expected prices to rise at the interest rate, known as the Hotelling rule[19]. If prices rise faster than the interest rate, it is efficient to store some of today's sales for future, higher value consumption. This shift to future sales increases today's price while depressing future prices, reducing the rate of price increase. Similarly, if prices rise more slowly than the interest rate, it is desirable to sell today rather than in the future. Overall, prices of storable goods like oil should not rise faster than the interest rate and only rise more slowly when little or no storage is taking place.

If market participants can correctly forecast the prices, an instantaneous auction will generally maximize the gains from trade, simply by arbitraging against future prices. Correct forecasts, however, are both computationally challenging and require data not usually available, like future supply and demand. Lavi and Nisan[20] present an algorithm with attractive worst-case performance in this setting.

Exchanges without money

Although the majority of exchange participants describe their preferences through financial bids and money changes hands with each transaction, there are cases when the exchange rules call for a pure trade, without money. In this case, the complexity lies in matching buyers with the sellers in an efficient manner.

Kidney exchanges, described by Roth *et al.*[21], are a primary example, where society frowns upon large sums of money changing hands. The difficulty in matching arises from the fact that a pure exchange is often not possible – someone who is willing to donate a kidney to a relative may not be a direct match. On the other hand, the donor may be a perfect match for someone on the list, but may hesitate donating to a stranger without a tangible benefit to his loved one. In this case a paired exchange, essentially forming a donation cycle with multiple operations happening simultaneously, can ameliorate the problem. An even better mechanism comes in a form of a list-exchange, where a donor provides a kidney to someone on the list in exchange for the list position being inherited by the donor's designate. This approach is efficient, both encouraging the donor to go through with the procedure, and respecting the list order subject to the compatibility constraints.

Sharing risk

There is a substantial conflict of interest between advertisers and publishers on advertising exchanges with regard to risk. Advertisers often want to pay per click – that is, only when the advertisement generates interest or observable activity. Publishers, in contrast, would prefer to be paid by the impression – that is, whenever the advertisement runs. Both have legitimate reasons, as some publishers produce little in the way of activity for advertisers and thus offer little value, and some advertisements do not generate clicks and thus under click-based pricing would be unlikely to produce value.

The conflict of interest associated with risk allocation is much greater than the risk-cost of a financial gamble. An advertiser buying millions of impressions (ads on web pages) with, say, a click probability of 0.1% faces very little financial risk, thanks to the law of large numbers. The real risk is that if the advertiser pays per the impression; the advertiser may wind up paying for a large number of 'junk' impressions. Similarly, if the publisher is paid per click, there may be very few clicks because, say, the ad run was not effective. Thus, the allocation of risk affects incentives. An advertiser who pays per impression has a strong incentive to create an ad that generates clicks, and a publisher who is paid per click has a strong incentive to generate good traffic to the website.

That the allocation of risk affects incentives is familiar from insurance, as insured parties have a weak incentive to take responsible care, a phenomenon known as moral hazard.

To resolve this inherent conflict, the exchange can take the risk, which is the strategy used by Google's AdSense. The risk could also be foisted on publishers, which is the strategy used by Yahoo!'s RMX. Or the risk could be imposed on advertisers, which is the strategy utilized by Google's display advertising exchange AdX. Finally, third parties could be used inside the exchange to absorb the risk, earning a profit in the process. Cavallo *et al.* (unpublished) consider the use of third-parties prediction in the exchange.

There are several principles associated with the allocation of risk. First, risk and incentives should, where possible, be allocated to parties that can do something about them. Thus, if the advertisers already have adequate incentives to create good ads, publishers should bear the risk, so as to give them an incentive to generate good traffic. Conversely, if traffic quality is easy to verify, the verification can be used to insure good behaviour by the publisher, in which case incentives can be pushed to the advertiser. Second, auditing is a substitute for incentives; where it is inexpensive to verify good behaviour, auditing can be used. Third, risk has a financial cost and is priced. Stocks with a higher risk produce a greater average return, and indeed the cost of higher risk has a market price. Ibbotson[22] provides an extensive list of the cost of capital by industry, which gives prices for various risk levels. Fourth, parties with deep pockets should bear financial risk, because the cost falls as the likelihood of a bankruptcy or other damaging outcome falls. Fifth, once a certain scale is reached, the costs are independent of who takes the risk. For example, the risk cost of, say, a $1,000 risk is the same for a million dollar company as a billion dollar company.

Tools

Even simple exchanges may be challenging for participants. Participants in advertising markets such as keyword search auctions offered by Google and Bing are faced with managing bidding on dozens, hundreds or in some cases millions of keywords. Tracking bids, prices, clicks and transactions that arise from them is a daunting task, made easier using automated tools. Moreover, nearly every buyer is in a similar position. Thus, it is advantageous to provide tools to participants to help bidders optimize their behaviour in the exchange. Market-provided tools accomplish two things: they help participants make accurate and effective decisions, boosting efficiency, and they save the cost of creating tools individually, which instead are amortized over many participants. Moreover, such tools can access market statistics and provide direct benchmarks.

In large, thriving markets, it is unnecessary for the exchange itself to provide the tools; third parties will spring up to do so. Many keyword bidders hire the services of search engine marketing firms, which provide and employ effective bidding and analytic tools.

A good balance is for the exchange itself to provide basic tools and leave difficult, sophisticated tools to third parties. Basic account management tools should be provided to all market participants, but it is important to appreciate that the provision of third-party tools serves the market itself and is a phenomenon to be promoted, not discouraged.

Third-party participation

Besides providing bidding and account management tools, third parties play two other important roles in exchanges. First, third parties may provide useful data. For example, third parties in advertising exchanges offer data on user interests, demographics and location that help advertisers optimize their campaigns. Targeting of advertisements can reduce the cost of a given level of impact by 75% or more. Price advice services for eBay may give buyers more confidence in purchasing. By improving buyer confidence, prices may actually rise.

In addition, in some settings, third parties may absorb risk. In particular, third parties can insure buyers, for example, letting sellers of advertising opportunity be paid for each showing of an ad (known as an impression) while the advertisers pay only if the ad is clicked.

Prediction and learning

Machine-learned signals are a large component in data-driven exchanges, such as modern advertising exchanges and stock exchanges. High-quality machine learning continually explores seemingly suboptimal strategies to mitigate the winner's curse and account for the dynamic nature of the system. This desire to optimize for the long term is at odds with the greedy strategy of taking the best available action at each time step that is best served by an auction mechanism.

An easy way to settle this conflict is to designate some transactions as 'learning events', where the participants agree to favour exploration over short-term gain. Although extremely simple, this approach is fraught with potential conflicts as the learning events are not independently distributed, and latecomers to the system greatly benefit from the learning done earlier. An alternative proposed by Li *et al.*[23] is to incorporate the long-term benefits of learning directly into the pricing created by the system, while maintaining short-term incentive compatibility for all parties.

Exchange earnings

Exchanges are costly to operate and participants should expect to cover their costs. Some marketplaces, like Apple's App Store that sells programs for the company's iPhone, may seek to make more money off the participants simply for the benefit of Apple shareholders. How should a marketplace make a profit?

The usual strategy is to charge a percentage, sometimes to buyers, sometimes to sellers, sometimes to both. The primary advantage of this strategy is simplicity. Sometimes the simple thing is the right thing to do, especially when the target profit level (and hence the percentage charge) is not very high. For small percentage levels, taking a small percentage of the transaction price is a simple, transparent thing to do.

As the target profit level rises, taking a percentage off the price creates an increasingly large inefficiency, often known as a dead-weight loss. In such cases economic analysis generally supports an alternate mechanism, the *value-added tax*. A value-added tax is typically a charge on the difference between the buyer's bid or offer price and the seller's asking price. A disadvantage of value-added taxes is that they distort incentives for the honest reporting of values, while an advantage is that they charge only when the exchange creates value. This is important because in many cases a seller has an alternative for the good or service, to use it himself or to sell in an alternate exchange. Large charges based on the price may prevent many transactions from taking place. However, a charge on the value generated by the exchange itself should not deter participation and eliminate fewer profitable transactions than a straight percentage of revenue.

Conclusion

A rich set of tools and know-how for building and improving exchanges is being developed. In particular, to summarize the findings, we recommend:

- Intentionally designing a language for expressing trades, that accommodates distinctions that matter substantially but not those of lesser importance.
- Designing the trading algorithm so that a straightforward strategy performs well.
- Permitting iterative adjustment of binding bid and asked prices, which simplifies participant strategies.
- Publishing suitably aggregated marketplace statistics, which makes markets more efficient in several different ways.
- Setting prices in a relatively coarse fashion without significant efficiency loss.
- Treating somewhat different products as identical to simplify participation.
- Keeping the exchange neutral, and not heavily tilting it toward one type of participant.
- Minimizing algorithmic complexity that makes sensible participation difficult.
- Creating rudimentary tools to help participants increase market efficiency.
- Attaining modest levels of revenue that can be raised with a straight percentage charge and switching to value-added pricing for greater levels of revenue.

Economic analysis is an imperfect guide to the design of markets. Consequently, new designs should be tested in a laboratory setting prior to use in the world. Moreover, an understanding of what participants actually value is critical to a successful design.

Notes

1. For surveys, see McAfee and McMillan[1], Milgrom[2,3], Krishna[4] and Klemperer[5]. Muthukrishnan[6] discusses some specific problems in ad exchange design.
2. Lu and McAfee[7] provide conditions under which an auction-based exchange out-competes an equally efficient bargaining-based exchange.
3. Milgrom[8,9] formulated this insight by 2007 but followed up with analysis of expressively complete languages. See Lahaie *et al.*[10] and the references therein for computationally feasible expressively powerful language design.
4. Facebook has recently begun running its advertising auctions using the VCG mechanism. VCG is also susceptible to collusion and may have very low revenue[16].

---

1. McAfee, R. P. and McMillan, J., Auctions and bidding. *J. Econ. Lit.*, 1987, **25**, 699–738.
2. Milgrom, P., Auctions and bidding: a primer. *J. Econ. Perspect.*, 1989, **3**, 3–22.
3. Milgrom, P., *Putting Auction Theory to Work*, Cambridge University Press, Cambridge, 2004.
4. Krishna, V., *Auction Theory*, Academic Press, 2002.
5. Klemperer, P., *Auctions: Theory and Practice*, Princeton University Press, Princeton, NJ, 2004.
6. Muthukrishnan, S., Ad exchanges: research issues. In Plenary talk at Workshop on Internet and Network Economics, 2009.
7. Lu, X. and McAfee, R. P., The evolutionary stability of auctions over bargaining. *Games Econ. Behav.*, 1996, **15**, 228–254.
8. Milgrom, P., Assignment messages and exchanges. *AEJ-Microecon.*, 2009, **1**, 95–113.
9. Milgrom, P., Critical issues in market design. *Econ. Inquiry*, 2011, **48**, 311–320.
10. Lahaie, S., Parkes, D. C. and Pennock, D. M., An expressive auction design for online display advertising. *AAAI*, 2008, 108–113.
11. Roth, A. E., Sonmez, T. and Utku Unver, M., The evolution of the labor market for medical interns and residents: a case study in game theory. *J. Polit. Econ.*, 1984, **92**, 991–1016.
12. Milgrom, P. and Weber, R., A theory of auctions and competitive bidding. *Econometrica*, 1982, **50**, 1089–1122.
13. Hendricks, K., Porter, R. H. and Boudreau, B., Information, returns and bidding behavior in OCS auctions: 1954–1969. *J. Ind. Econ.*, 1987, **35**, 517–542.
14. Cramton, P. and Schwartz, J., Collusive bidding in the FCC spectrum auctions. *Contrib. Econ. Anal. Pol.*, 2002, **1**, 1.
15. Halevy, A., Peter, N. and Pereira, F., The unreasonable effectiveness of data. *IEEE Intell. Syst.*, 2009, **24**, 8–12.
16. Ausubel, L. and Milgrom, P., The lovely but lonely Vickrey auction. In *Combinatorial Auctions* (eds Cramton, P., Steinberg, R. and Shoham, Y.), MIT Press, 2005.
17. Brewer, P. J. and Plott, C. H., A binary conflict ascending price (BICAP) mechanism for the decentralized allocation of the right to use railroad tracks. *Int. J. Ind. Organ.*, 1996, **14**, 857–886.
18. Klemperer, P., The product mix auction: a new auction design for differentiated goods. *J. Eur. Econ. Assoc.*, 2010, **8**, 526–536.
19. Hotelling, H., The economics of exhaustible resources. *J. Pol. Econ.*, 1931, **39**, 137–175.
20. Lavi, R. and Nisan, N., Competitive analysis of incentive compatible on-line auctions. *Theor. Comput. Sci.*, 2004, **310**, 159–180.
21. Roth, A. E., Sonmez, T. and Utku Unver, M., Kidney exchange. *Q. J. Econ.*, 2004, **119**, 457–488.
22. Ibbotson Associates, *Ibbotson Cost of Capital Yearbook 2011*, Morningstar, Inc.
23. Sai-Ming, Li., Mahdian, M. and McAfee, R. P., Value of learning in sponsored search auctions. In Proceedings of the 6th International Workshop on Internet and Network Economics, 2010.