Behavioral Efficiency: Definition, Methodology, Demonstrations

Ronald M. Harstad*

Economics, University of Missouri

harstad@missouri.edu

February 2012

Abstract

Laboratory experiments employing an induced-values methodology report on allocative efficiencies observed. That methodology requires experimenters know subjects' motivations, impossible in field experiments. Allocative efficiency implies a hypothetical costless aftermarket would be inactive. An allocation mechanism's outcome is defined to be *behaviorally efficient* if an appropriate aftermarket is actually appended to the mechanism and at most a negligible size of remaining mutually beneficial gains identified. Methodological requirements for behavioral efficiency observation are provided. A first demonstration observes significantly greater behavioral inefficiencies in second- than in first-price auctions. A simple field demonstration indicates when a public good increase can be observed to mutually beneficially cover marginal cost, without knowing valuations. Several empirical issues that arise are noted.

C9; C93; D01; D61; D03; D46; Keywords: behavioral efficiency, field experiment methodology, allocative efficiency, revelation of valuations, aftermarkets

---

* I thank, without implicating, the Economic Science Laboratory at the University of Arizona and LAMETA at Universite de Montpellier for providing facilities, Cathleen Johnson and Marc Willinger for organizing experimental facilities and arrangements, Eric Cardella, Raymond Chiu and Dimitri Dubois for assistance in conducting experiments, Katsunori Yamada for assistance in data analysis, Dubois and Rafaele Preget for translating instructions, Willinger and Sophie Thoyer for organizing financial support, David Levine, Alvin Roth, Larry Samuelson and scores of seminar attendees for useful suggestions, the J Rhoads Foster Professorship Endowment of the University of Missouri, the Freshwater Group Research Fund of the University of Arizona, CNRS and INRA for financial support of the experiments, and ISER at Osaka University for support during the writing.

1. Introduction

Recommendations for policy adoption or alteration are more valuable if evidence of the size of shortfalls from allocative efficiency can be provided for the allocation mechanisms or policy instruments under consideration. Such evidence has so far come from laboratory experiments using an *induced-values* methodology for, e.g., an abstract commodity.[1] That methodology requires that experimental subjects' motivations are known to the experimenter, and as such is unavailable for field experiments.

Field experiment transactions might involve (usually subsidized) provision and/or allocation of: irrigation water, adult education, childcare, pollution permits, microfinance, insurance against background risk, or similar "naturally occurring" goods, services and contracts. In "conducted" field experiments, these transactions occur on a market constructed and controlled by the experimenter; in "natural" field experiments, the experimenter observes but cannot control a naturally occurring market.[2] The experimenter can observe transacting behavior but not valuations or motivations, and so cannot calculate the Pareto set nor measure shortfalls from it.

This paper proposes a definition and a specific but broadly usable methodology to allow observations relevant to allocative efficiency in field experiments, which observe only behaviors that stem from unobserved motivations and preferences (and

---

[1] A subject j might, for example, be told that she is one of *N* buyers or *M* sellers of an abstract good called *X* (not "tennis lessons" or "coupons for video downloads") that will be traded, with her payoff being the difference between trading prices and induced values, as in "the first unit of *X* you buy can be resold to the experimenter for $8.75, the second for $6.80, the third for $5.10" to a potential buyer, or "the first unit of *X* you sell can be obtained from the experimenter for $3.10, etc." Cf. e.g., Smith [1962], [1976], Davis and Holt [1993], Kagel and Roth [1995], Plott and Smith [2008], and original sources cited in the latter two works.

[2] I am defining "natural" field experiments somewhat differently than Harrison and List [2004], as the issue of experimenter control seems more critical than whether subjects are aware they are in an experiment.

sometimes even incompletely observed feasibility constraints). An appropriately designed aftermarket is appended to the experiment. Theoretical issues arising with this methodology are described. Following that, sections 5-11 provide a first demonstration of the concept, a simple but concrete example of the appropriate usage of a properly constructed aftermarket to observe allocative efficiencies (or shortfalls therefrom) without relying on knowing subjects' motivations. It finds first-price auctions less behaviorally inefficient than second-price auctions, and measures efficiency shortfalls; in this context, subjects' bidding was unaffected by knowing there would be an aftermarket.

Then, a first field aftermarket is reported in section 12, observing whether an increase in output of a public good from an ad hoc starting point can be achieved as a mutually beneficial reallocation. Finally, several empirical issues that arise in the consideration of this methodology are discussed.

Among the finest examples of how far field experiments have been able to go in the direction of inferring efficiency conclusions from observations is Bohm [1984]. Peter Bohm convinced the Swedish federal government to let him control whether an indivisible public good would be produced or not, an office that would collect and provide certain statistical data to local governments. Randomly splitting the local governments into two groups, he announced rules to one group that gave them an incentive to understate willingness-to-pay (WTP), and to the other an incentive to overstate WTP. The observed sum of stated WTPs slightly exceeded cost, and mean WTP per capita was not markedly different between the two groups. Nothing Bohm could observe, however, would let him infer whether providing the statistical office would yield an efficiency gain (no subject had an incentive to correctly state WTP).[3]

---

[3] Bohm believed the binary nature of the public good decision he studied was a large advantage: "The case of divisible public goods, requiring the revelation of WTP functions, or at least WTP for several alternative quantities, is referred to the science fiction department for the time being." (pp. 138-9) This paper, decades later, offers an avenue toward science fact.

1.1. Pareto Efficiency Reinterpreted

In an economy with I individuals and C commodities, let x_i be the C-dimensional allocation to individual i, u_i his utility function, and $x = (x_1, \ldots, x_I)$. Then the usual definition of a Pareto-efficient allocation is that it satisfies

(P): $\text{Max}_x\ u_1(x)$, subject to: $u_j(x) \geq \underline{u}_j$, all $j = 2,\ldots, I$, and to feasibility conditions.

In the 1950's it became commonplace among several developers of general equilibrium theory to add imagery in a reinterpretation of this maximization problem: Suppose an allocation were to be Pareto-efficient. Then a hypothetical costless aftermarket would be *inactive*, for the simple reason that, upon reaching a Pareto-efficient allocation, there would be no remaining mutually beneficial transactions to exploit.

This reinterpretation is informationally dissimilar: the maximization in (P) is clearly tied to knowledge of motivations and valuations (u_i, after all), while the counterfactual aftermarket is tied to hypothetical transactions (that is, to hypothetical behaviors). In principle, as transactions can be observed in field experiments, they might avail themselves of aftermarkets.

2. A Definition

So I define *behavioral efficiency*: an outcome of an allocation mechanism is said to be *behaviorally efficient* if an appropriate (incentive-compatible, suitably transparent, and approximately costless) aftermarket is actually (and immediately) appended to the allocation mechanism[4] and at most a negligible aggregate size of mutually beneficial gains is observed on the aftermarket. Natural extensions of the definition include at least the following: [a] an allocation mechanism is said to be behaviorally efficient in a particular context if it reliably yields behaviorally efficient outcomes; [b] a social or

---

[4] Any method, however informal, of reaching an allocation is herein labeled an allocation mechanism.

economic policy *Y* is said to be *behaviorally less inefficient* in a particular context than an alternative policy Z if the shortfall from behaviorally efficient outcomes under policy *Y* is robustly observed in such aftermarkets to be significantly smaller than observed under *Z*.

3. Aftermarket Methodology

The aftermarket referred to in this definition must be designed and implemented so as to support the intended normative interpretation. It likely aids first to set aside straightforward disqualifications: [i] In general, simply repeating an allocation mechanism does not suffice to draw meaningful conclusions about efficiency of the initial application of the mechanism (an illustration is in section 4).[5] [ii] Whatever its structure, a *resale* market (Zheng [2002]) does not suffice. A key terminological distinction: unlike resale markets (e.g., for US Treasury debt), an *aftermarket* necessarily involves *the same* economic actors as the original market (original allocation mechanism), none added and none absent.[6] Imagine $100K in 5-year T-notes is sold today, by their purchaser at a Treasury auction on the third Tuesday of the month before last, to a regional bank that did not compete in that particular Treasury auction. Today's resale in no way implies an inefficiency in the allocation that resulted from that auction. [iii] Later transactions involving an informationally distinct commodity cannot support interpretation of an earlier allocation as inefficient. For example, suppose one of a group of competing used-car dealers obtains a particular car at an auction of cars whose leases have ended. Some days after the auction, the consigner of this particular car agrees to allow the winning bidder to return the car and be given a full refund; that winning bidder continues to be considered a financially reliable bidder by the auctioneer. Even if exactly the same

---

[5] Were repeating the same mechanism to suffice in some special circumstance, likely it would create needless confusion for subjects.

[6] This definition is clearly implied in the half-century-old reinterpretation of Pareto efficiency noted in section 1.1.

set of bidders are competing when the car is re-auctioned, a different bidder winning the re-auction does not imply any inefficiency of the original auction. The knowledge that the car was returned, inferred from the fact of its re-auction, leads to a realization that a quality issue unsuspected as of the original auction has since surfaced, thus to a different commodity being allocated at re-auction.

To shed light on behavioral efficiency, the aftermarket must be constructed so as to *identify* any and all remaining mutually beneficial transactions involving the same set of traders, under the same information as occurs when the original market (or other mechanism) reaches an allocation.[7] This requires *revelation*: that subjects' behavior in the aftermarket can be interpreted as revealing the border between potential transactions they prefer to make and prefer not to make, thus as revealing all relevant willingnesses-to-pay and willingnesses-to-accept.[8] In straightforward situations, this can be accomplished via a typical incentive-compatibility characterization: that the price to any partner in any transaction be independent of his or her own behavior, with the impact of the behavior limited to affecting whether (to be precise, the probability with which) an aftermarket transaction occurs.

As formalized in the next section, the aftermarket should be designed in such a way that a posited equilibrium behavior in the original allocation mechanism,

---

[7] I follow a century-old tradition in welfare economics that externalities to a transaction are either explicitly modeled or ignored. Thus, if allocations *A* and *B* differ only in that in *A*, a cocktail dress remains in company C's inventory, while in *B*, Ginger buys the dress from company C at a mutually beneficial price, then allocation *B* is considered Pareto-superior to *A*. This ignores the possibility that Ginger might later attend a party where seeing the new dress makes Rosemary less happy with her wardrobe. In contrast, where a field study is focused on externalities, the aftermarket must observe all possible incremental utilities or disutilities resulting from recontracting, not just the transactors' evaluations.

[8] I know of only two antecedent reports of experiments employing aftermarkets, Grether, Isaac and Plott [1979, 1989] and Rassenti, Smith and Bulfin [1982]. Both experiments are fine laboratory studies of airport landing rights; in neither is the aftermarket designed to identify all possible remaining gains from trade, and in both the induced values are utilized to analyze original and post-aftermarket efficiency.

together with truthful revelation in the aftermarket, constitute an equilibrium in the mechanism-cum-aftermarket game.

Importantly, this is *not* because equilibrium behavior is assumed or even expected in either the original allocation mechanism *or* in the aftermarket. Rather, it is because an aftermarket not so stringently designed could not possibly be informative as to the original outcome's efficiency.

If the aftermarket had to be an allocation mechanism in its own right, this revelation requirement would typically be impossible (Myerson and Satterthwaite [1983] provide an impossibility theorem for perhaps the simplest case). However, the aftermarket is to be appended to a mechanism, not to obtain an allocation, but merely to normatively categorize the allocation that was reached *before* the aftermarket was used. The simplification thus obtained is characterized in section 4.

This bears emphasis: for the purpose of behavioral efficiency interpretation, an aftermarket *does not* have to be an allocation mechanism, and the behavior preceding the aftermarket *does not* have to be equilibrium or even rational behavior.

That the aftermarket be approximately free of transactions costs, and that it be suitably transparent, necessarily have less exacting interpretations; these are arenas where employment of aftermarkets shifts from science, narrowly construed, in the direction of art. Negligibility of transactions costs is most usefully evaluated relative to the size of potential mutual gains from further transacting. Indeed, transactions costs yield a calibration: an appropriate aftermarket identifies all mutually beneficial transactions for which the perceived gains from trade exceed the perceived transactions costs.

When subjects have already been congregated, either physically or via simultaneous interaction on the Internet, an aftermarket run fairly quickly and with

simple, transparent tasks for subjects is likely to sluff off transactions-costs concerns.[9]

Experimental psychology and laboratory experimental economic literatures yield insights into transparency that are extensive, although often anecdotal and always subjective.[10]

3.1. Required Nature of Potential Aftermarket Transactions

A useful aftermarket needs to observe valuations with respect to any potential transaction that may be mutually beneficial; this need will vary with the topics and mechanisms of field studies. When allocation of identical units of a single private good is the concern (or when identical units of multiple goods are at stake, but issues of complementarities or income effects can reasonably be assumed absent), observations of behavioral valuations of bilateral trades suffice. When goods L and R are complements for some subjects, the possibility that subject 1 might value an additional set $\{L, R\}$ by an amount sufficient to compensate both subject 2 for forgoing one unit of L and subject 3 for forgoing one unit of R must be observable by design.

Another important example arises when the original allocation mechanism determines a quantity of a public good to be produced, and an allocation of its production cost. Now an aftermarket merely observing potential bilateral transactions is insufficient; transactions which alter public good output and attain some adjustment of cost shares in accordance with increased or reduced public good production cost must be considered, as seen in section 12.

---

[9] When congregating is only required for the aftermarket, it may be that an appropriate aftermarket design compensates subjects for the costs of congregating, being careful to compensate in a manner unrelated to observed aftermarket activity.

[10] That price-clock-based ascending-price mechanisms are notably more transparent than sealed-bid mechanisms seems a reasonable inference to draw from the laboratory experiments reported in Harstad [2000]. For this reason, such ascending-price mechanisms are used both in the lab demonstration of sections 5-11 and in section 12's field demonstration.

Some field studies will require the aftermarket be designed so as to observe valuations of dynamically structured contracts. Allocations of common-pool resources are examples.

The usual characterization of efficiency via (*P*) above determines marginal conditions and assumes the appropriate convexities to imply that a local optimum is a global optimum. A similar limitation may be needed in many cases to keep aftermarkets sufficiently simple and straightforward. For example, consider examining in an aftermarket both an increase and a decrease in public good output, each by some more-than-differential amount that is in context small. Concluding a behaviorally efficient outcome from an observed inability to find any mutually beneficial increase or decrease by that given amount assumes the unobserved motivations were consistent with marginal valuations decreasing more rapidly than marginal production cost. Should a behavioral inefficiency be found, the study would indicate in which direction public good output could be altered so as to obtain a perceived mutual gain, but not how far such a movement could continue. In most contexts, the imaginable alternative of checking several possible increases in public-good output of varying sizes, and corresponding decreases, is likely to rob an aftermarket of a required low-transaction-costs character.[11]

[11] Correspondingly, suppose an aftermarket appended to a mechanism allocating a given quantity of identical units of a private good were to observe that the potential buyer willing to pay the most for an additional unit could not cover the lowest price at which some potential seller was willing to provide the additional unit. Assuming that there was no mutually beneficial trade in which this buyer would acquire two units of the good (thus, assuming unobserved motivations included diminishing marginal utility) might be preferable to running an aftermarket that priced 2-unit (and perhaps 3-unit) trades as well as 1-unit trades, and allowed for multiunit trades to have multiple parties on the same side of the trade.

4. Theoretical Issues Raised

Suppose the allocation mechanism being studied is sufficiently formal to permit analysis as a game *G*.[12] Let *G*+ be the strategic-form game with nature that implements a properly constructed aftermarket, where nature is modeled in *G*+ as making all choices that nature or any other player made in *G*; let **G** = {***G***, *G*+}. Then proper construction for the aftermarket to yield behavioral efficiency observations requires that truthful revelation for all real players be a dominant strategy in *G*+ *for any* play in *G*, including any equilibrium *E* of *G*.

Otherwise, a behavioral efficiency (or inefficiency) conclusion is unwarranted. It bears repeating: this is *not* because equilibrium behavior in *G* is assumed or even expected. [i] If the design of *G*+ distorted incentives so that *E* was no longer an equilibrium of *G* when **G** appended *G*+, then even rational behavior following which the aftermarket measures inefficiencies (behavior under incentives other than those generated by *G* itself) *cannot possibly* be the behavior we desire to measure (that of *G* itself). [ii] If truthful revelation in *G*+ were not an equilibrium in *G*+, following the play of *E* in *G*, then nothing could be inferred from the observed behavior in *G*+ about the efficiency of behavior in *G*. [iii] Of greatest import is the sequential-rationality requirement: that following any behavior in *G*, whether equilibrium or not, it remains a dominant strategy to truthfully reveal in *G*+. Only then do we have the *possibility* of using aftermarket behavior to measure inefficiencies in *G*.

Aftermarket construction can thus be viewed as a particular type of mechanism design problem. While formal constraints of mechanism design are often limiting, the mechanism design challenge posed here should always be attainable. An aftermarket constructor has two critical dimensions of flexibility generally unavailable in mechanism design: the field experiment [a] does not have to balance the budget,

---

[12] This supposition is not trivial: some of the cultural incentive schemes discussed in Ostrom [1998] may be difficult to formalize as allocation mechanisms.

though hopefully limiting the size of any deficit; [b] does not have to implement any transactions observed to be mutually beneficial with probability one, but merely with positive probability.[13]

As a simple illustration, suppose a field experiment has observed a failure to reach an agreeable transaction in a bilateral bargaining situation. As motivations are unobserved, it is unknown whether the failure was an efficient outcome or a mutually beneficial bargain was possible. Myerson and Satterthwaite [1983] demonstrate that no mechanism can insure efficient outcomes when the potential seller's and potential buyer's valuations are private information. However, at least three distinct aftermarket constructions can observe whether the outcome was behaviorally efficient.[14] Each asks the seller to state the lowest price that he is willing to accept, and the buyer to state the highest price that she is willing to pay.[15] The experimenter has carefully explained in advance to the subjects what use will be made of their responses. Aftermarket version 1 will implement the transaction whenever B, her stated willingness-to-pay, exceeds A, his stated willingness-to-accept, with the pre-announced rule that she will pay A, and he will receive B. Version 1 is incentive-compatible, implements any transaction observed to be mutually beneficial, and requires the experimenter to cover the deficit $B - A$. Aftermarket version 2 draws a random variable R from a distribution exogenous to all information provided by this pair of subjects (perhaps uniform on [0.25 W, 1.75 W], where W is a publicly available average price from a prior survey of similar transactions in the economy), and transacts at random price R if $B \geq R \geq A$.

---

[13] It may be worth noting that, where the implementation is financial, this implies a positive probability that the commitments made are financially incurred. I see no opportunity for surveys about whether subjects wished to reallocate, or about hypothetical terms under which subjects wished to reallocate, to substitute for an aftermarket.

[14] If the failure to reach a transaction occurred in a natural field experiment, the aftermarket would require a transition to a conducted field experiment.

[15] Depending on context and the background and culture of the subjects, this may well *not* be the language in which the experimental instructions state the request.

Aftermarket version 2 is incentive-compatible, implements any transaction observed to be mutually beneficial *with positive probability*, and balances the budget. Aftermarket version 3 also draws a random variable R from an exogenous distribution, transacts if $B \geq R \geq A$, but she pays R and he receives 1.05 R, achieving incentive compatibility, implementing with positive probability transactions observed to be mutually beneficial, but requiring the experimenter cover a deficit of 0.05 R when transactions occur.

Note that mechanism design requirements can still impinge on experimental desiderata. In particular, consider an aftermarket construction which attempted to alter aftermarket 1 above by only transacting when the deficit $B - A$ did not exceed a maximum desired experimenter cash infusion M. This construction would no longer suffice for incentive compatibility: it is possible that the seller would attain the outcome of no transaction and no gain if he truthfully stated A, as $B - A$ might exceed M, while some overstatement $X > A$ might yield a gain of $B - X > 0$ should $B - X$ be less than M. Thus, instead of an incentive to truthfully reveal his willingness-to-accept, the seller (and the buyer) would optimally trade off a lower gain in the event of transaction against a higher probability of a gain by some degree of misstatement. (Even if, instead of announcing M, the experimenter merely stated that the transaction would occur "unless the deficit were too large," the construction would still be insufficient to warrant conclusions as to behavioral efficiency.)

5. The Setting of the Initial Lab Demonstration

Five-bidder sealed-bid auctions of a single abstract asset were conducted, in seven sessions (110 subjects) via first-price rules, and in six sessions (85 subjects) via second-price rules.[16] There were generally 10, 15, 20, 25 or 30 subjects in the

[16] Thus, comparisons across pricing rules are between-subject comparisons.

laboratory during a session, with random reassignments into groups of five each period.[17]

Affiliated asset valuations (Milgrom and Weber [1982]) for subjects were determined as follows. In each period, first a random number C, called a *central tendency*, was drawn uniformly from [\$50, \$1000] (all random variables are multiples of \$0.01). Then, given a realization c of C, for each subject j an *estimate* X_j was drawn uniformly from [\$(c – 10), \$(c + 10)], conditionally independent. Finally, *asset value* to subject j was $V_j = (3/4)X_j + (1/4)C$; this system incorporates private values (the first term, X_j) to introduce efficiency issues, as well as a natural, small common-value component (C).[18] These rules were carefully explained and examples given. C was not revealed to subjects until end-of-period feedback, which gave a complete, anonymous report of C, X_j's, V_j's, all behavior, and profit calculations. The instructions stated that a reserve price, below which the asset would not be sold, was drawn anew before each auction, uniformly from [\$(c – 10), \$(c – 6)], and would not be revealed until end-of-period feedback.[19]

Subjects began the experiment with a bank balance of \$12, with profits added and losses subtracted during the session, and the final balance paid in cash. These valuation procedures call for a small winner's curse correction; the 90% confidence interval for the loss in the event a winning bid exceeded the symmetric, risk-neutral equilibrium bid by exactly the winner's curse correction (were all rival bids in

---

[17] Subjects were University of Arizona undergraduates, recruited campuswide via website, and sat at visually isolated computers. A second-price auction session for which less than ten subjects showed was eliminated from data analysis. The experiments were conducted in October and November 2009, using the Z-Tree programming environment (Fischbacher [2007]).

[18] A principal reason for including a common-value component was to avoid a throw-away bid problem: with independent private values, most values will yield so low a chance of winning as to make serious consideration of what to bid not worthwhile. Unfortunately, how high a value a subject has to draw before he or she chooses to pay attention is unobservable. In the current design, all estimates between \$60 and \$990 have the same expected profitability, removing throwaway bid concerns. (Data analysis only includes the cases, almost 99%, where subjects' estimates were in the [\$60, \$990] range.)

[19] As expected, the reserve price was never binding.

equilibrium) is about [$1.50, $4.25]. Thus, three to four such losses could likely be handled without the balance going negative.[20]

6. The Aftermarket Methodology Implementation

To discern from subjects' behavior whether an auction attained an efficient outcome, the experiment appended an aftermarket designed as follows. Once all subjects had typed and submitted their bids, the winning bidder was determined (throughout by fair random tie breaking if necessary). Then each bidder was informed of the price determined in the auction and whether his bid acquired or did not acquire the asset. Some seconds later, the aftermarket was begun; the subjects had been told before bidding that this aftermarket would follow the auction.

A price clock ticked up on all subjects' screens, rising by $0.25 every two seconds (though more slowly in the first period with an aftermarket), beginning at a random price calculated to be acceptable to all subjects but noisy enough to avoid revealing information about the still-unknown *C*. The bidder who acquired the asset was labeled the *offerer*, and asked to click the "Accept" button on the screen when the price reached the lowest price at which he was willing to sell the asset just acquired in the auction to one of the four rival bidders. Each of the four bidders that did not submit the highest bid was asked to do nothing so long as the prices being shown were prices at which he would be willing to buy the asset from the offerer, and then to click "Accept" at the highest such price. No subject observed any information about other subjects' behavior in the aftermarket until all five had clicked on a price.

Instructions had carefully described the rules relating these Accept Bids (of the four bidders who did not acquire the asset) and Accept Ask (of the offerer) to possible aftermarket transactions. [1] If the offerer's Accept Ask exceeds all four Accept Bids, there is no aftermarket transaction. [2] If at least two Accept Bids are

[20] If a subject's balance became negative, he was given a $20 loan to be repaid out of his final bank balance. Two of 195 subjects could not quite repay the loan; it was of course forgiven and they were paid only the usual $5 show-up fee.

no lower than the Accept Ask, the asset is transferred from the offerer to the bidder selecting the highest Accept Bid, at a price set by the second-highest Accept Bid. [3] If the highest Accept Bid exceeds the Accept Ask and it exceeds all other Accept Bids, a random number *R*, drawn before the auction from multiples of \$0.01 in [\$*c*, \$(*c* + 15)] equiprobably, determines the aftermarket outcome. If R falls between the highest Accept Bid and the Accept Ask, the asset is transferred from the offerer to the bidder selecting the highest Accept Bid, at a price set equal to *R*; otherwise, there is no aftermarket transaction.

This aftermarket design justifies the inferences about behavioral efficiency to be drawn from observations of aftermarket behavior: any mutually beneficial trade revealed is transacted with positive probability, and in no aftermarket transaction is the price determined by the behavior of either transacting party. Among possible aftermarket designs, prior experimental evidence (Harstad [2000]) suggests the use of the price clock makes aftermarket incentives as transparent as possible. Whenever at least one nonacquiring bidder selects an Accept Bid above the offerer's Accept Ask, a mutually beneficial trade that the auction did not achieve has been identified (whether or not the aftermarket actually transacts that trade).

7. Session Protocol

Each experimental session ran 150 minutes and followed a multi-phase protocol, to build the desired treatment step-by-step from simpler games. After instructions regarding the whole session and the first phase, that first phase exposed subjects to the software of the aftermarket, without introducing the word "aftermarket." In phase 1 (4-5 periods), each subject was informed of a list of all five private values of the abstract asset (told which was his value), which were drawn i.i.d. uniform on [\$5, \$10]. Per instructions, one subject was chosen at random to be the offerer, the

others bidders.[21] As just described, the offerer was asked to click on an Accept Ask, the four bidders to click on Accept Bids. Then the aftermarket rules above were used to determine payoffs for the period, which were simply asset value minus transaction price for the buyer, and transaction price minus asset value for the seller, if there was a transaction, and zero for all non-transacting subjects.

Further instructions were distributed and read before each following phase. Phase 2 (6-7 periods) introduced private information, with subjects' private values first revealed to all group members (anonymously) during end-of-period feedback. Phase 3 (6-7 periods) introduced two changes: [i] all five subjects were now bidders asked to select Accept Bids (that is, in a closed-clock variant of an English auction), and [ii] the private values were now affiliated (as in section 2, except that $V_j = X_j$). Phase 4 (6-7 periods) set aside the price clock, introducing bidding in a sealed-bid auction (first- or second-price, depending on the session). Phase 5 (8-11 periods) introduced affiliated values, via $V_j = (3/4)X_j + (1/4)C$ as in section 5.

All this led to the phase of principal interest, phase 6, which re-introduced the software from the first two phases, but with the offerer being the bidder who acquired the asset in the sealed-bid auction, and the following price-clock activity called an aftermarket. Phase 6 was generally limited by the time constraint, 6-11 periods. The session ran faster when there were fewer groups (with the software always waiting for the last subject in the session to bid, to peruse feedback, etc.); in four of the first-price sessions, we were able to run a final phase 7. Phase 7 had aftermarkets only in even-numbered periods, with the sealed-bid auction the final determination of period profits in odd-numbered periods. In the other nine sessions, phase 6 was the final phase.

[21] To generate possible gains from trade frequently, the program chose the offerer from the highest, second-highest, …, lowest private values with probabilities {1/8, 1/8, ¼, ¼, ¼}.

8. Contrasting Predictions

Auction theory predicts aftermarket activity with positive probability following first-price auctions, but with zero probability following second-price auctions. It is straightforward to show that the unique, risk-neutral, symmetric Bayesian equilibrium of auction-cum-aftermarket (either first- or second-price) is to submit one's equilibrium bid in the auction and to select an Accept Bid or Ask in the aftermarket most nearly equal to one's rational Bayesian-updated willingness-to-pay or -accept. In this equilibrium, publicly announcing the price attained in a first-price auction informs each losing bidder (but not the winner) of the amount by which his bid lost. Whenever a bidder lost by a sufficiently small margin, rational updating leads to his willingness-to-pay exceeding the winning bidder's willingness-to-accept (as he knows of a second estimate nearly as high as the winning bidder's estimate, which can be inferred from the price set by the winner's monotonic equilibrium bid function).

No similar occurrence is possible following announcement of the price in second-price auctions. Here the price reveals the private information of the second-highest bidder, who Bayesian updates on the basis of learning that one rival estimate was higher and three lower, and this leads to a willingness-to-pay that exceeds his equilibrium bid, while the winning bidder's updating leads to a willingness-to-accept that is less than his equilibrium bid. However, the second-price auction equilibrium is envy-free: these two adjustments of willingness-to-pay and to accept sum to less than the difference between the two highest bids, and thus do not change their ordinal rank.[22]

To my knowledge, prior auction experiments have either induced private values (independent, as in phases 1 and 2, or affiliated, as in phases 3 and 4) or common

[22] Though stated slightly differently, the results in this and the previous paragraph are not new, and can be pieced together from Milgrom [1981], Milgrom and Weber [1982], and Harstad and Bordley [1996].

values (modifying section 2 so that $V_j = C$, hence there is no inefficiency generated should the bidder with the highest estimate be outbid). Nonetheless, in both settings, bidders have bid significantly above the risk-neutral symmetric Bayesian equilibrium (Kagel [1995]), and (more pertinent here) have exhibited more heterogeneity in this overbidding in second-price than in first-price auctions.[23] Thus, prior laboratory experiment results predict more aftermarket activity following second-price auctions.

9. Aftermarket Observations

First-price [second-price] auctions were observed to be behaviorally efficient in 72% [57%] of the observations (cf. Table 1). In 28% of 203 first-price auctions, and 43% of 142 second-price auctions, at least one bidder who was outbid was observed to be willing to buy the asset from the high bidder for mutual gain. (These percentages naturally sum occurrences where the aftermarket transacted with those where the random price fell below the high bidder's Accept Ask or above the one Accept Bid exceeding that Accept Ask.) This difference is significant at the 1% level in a Pearson test.

Where aftermarket behavior exhibited such gains, the difference between the most an outbid bidder will pay and the least the high bidder will accept is a behavioral measure of the shortfall from efficiency, averaged in row 3. While shortfalls when observed were larger in first-price auctions, when zero shortfalls are averaged in for the behaviorally efficient outcomes, the expected shortfall in row 6 is smaller for first-price auctions.

It bears emphasis that, while these auctions sold induced-value assets, the behavioral efficiency and shortfall measures make no use of any information

[23] Kagel and Levin [1986] report on 199 first-price, common-value auctions, and Kagel, Levin and Harstad [1995] on 154 second-price, common-value auctions. To adjust for varying number of bidders, I calculated a statistic for each session that takes the frequency with which the high signal holder was the high bidder and subtracts $1/n$. The weighted (by number of auctions) average of these statistics was 50.93 for first-price auctions and 38.73 for second-price auctions.

contained in the induced values. These reports stem solely from subjects' behaviors: their sealed bids, Accept Bids and Accept Asks, and in no way depend on any information about subjects' motivations.

		Table 1	
		Observations	
		First-Price Auctions	Second-Price Auctions
1	Aftermarkets Observed	203	142
2	Behaviorally Efficient	72%	57%
3	Mean of Shortfalls	$4.45	$4.16
4	Shortfall Capacity		$10.26
5	Aftermarket Fraction		41%
6	Expected Shortfall	$1.24	$1.79

Reports of allocations reached in induced-values experiments can provide efficiency measures in percentages, because the dollar value of total gains from trade in Pareto-efficient allocations can be calculated from the induced values. This methodology cannot be used in the field.

In some situations, behavior in the original allocation mechanism can offer a benchmark for the economic significance of the size of shortfalls from efficiency. This experiment demonstrates both the possibility and its limitations.

Second-price auctions are incentive compatible, in that the bid selected determines only whether the bidder wins or not; the price is solely determined by the highest rival bid. In particular, the risk-neutral symmetric equilibrium bid is the expected asset value conditioned on an assumption that the bid is pivotal.[24] For the

[24] For private values, this is the dominant strategy discovered by Vickrey [1961]. This feature of second-price, common-value auctions was first found by Matthews [1977]; the intuition is presented in Harstad and Bordley [1996].

distributions of section 2, this implies that bids should differ from bidders' expected values by a constant.[25] Hence, differences between two bids submitted in second-price auctions should (i.e., in symmetric equilibrium) be equal to the differences between the two bidders' willingnesses-to-pay and therefore measure the gain from trade if the asset were hypothetically to be transferred from the lower bidder to the higher bidder. These observations include 586 differences between a losing second-price bid and the high bid in the same auction. The average of those 586 bid differences is the $10.26 reported in row 4 above. This is the *capacity* for shortfalls from efficiency in the sense that, had the original auctions allocated the asset equiprobably among inefficient acquirers, aftermarkets that reallocated to the efficient acquirer could average $10.26 in gains from trade unattained by such a complete misallocation to a random inefficient acquirer.

The best measure I can envision to attach an economic significance to the $4.16 mean of shortfalls revealed by aftermarkets following second-price auctions is that it is 41% of the shortfall capacity.

The absence in Table 1 of a comparable benchmark for first-price auctions is not an oversight. Differences between a losing first-price bid and the high bid in the same auction could be averaged. However, without the incentive compatibility of second-price auctions, these first-price bid differences have no similarly strong argument to measure gains from hypothetical transfers between the bidders: optimizing the {profitability given winning/probability of winning} tradeoff in the risk-neutral symmetric equilibrium of the first-price auction yields a nonlinear term in the bid function (corresponding to the term that is a constant in second-price auctions). In that equilibrium, if bidder *A* outbids bidder *B* by $8, *A*'s willingness-to-pay exceeds *B*'s, but the $8 bid difference is not a measure of the willingness-to-pay difference.

[25] This neglects estimates in the ranges [$50, $60] and [$990, $1000], for which the difference is not constant.

10. Submitted-Bid Impact

The unique risk-neutral symmetric Bayesian equilibrium of the game consisting of the sealed-bid auction (either pricing rule) followed by the aftermarket is for each bidder to make the same equilibrium bid as if there were no aftermarket, and then truthfully reveal in the aftermarket. Despite the theory, it is an empirical question whether subjects bid the same way when they know an aftermarket will follow; there might be reasons subjects would find for bidding less, or for bidding more, in an auction when knowing there will be an aftermarket.[26] The protocol in section 7 is designed to shed light on this question.

The following linear bid function was estimated separately from the first-price and second-price data:

$$M_{st} = \text{const} + \beta_x \, \text{Exper}_{st} + \beta_a \, \text{After}_t + \text{error}_{st},$$

where the markup M_{st} was the observed bid minus the asset value estimate X_{st} for subject *s* in period *t*; Exper_{st} was a control for possible learning effects, the number of periods of experience in the affiliated-values auctions; After_t was a dummy variable taking the value 1 if the subject knew the auction in period *t* would be followed with an aftermarket, 0 if the subject knew the auction would not be followed with an aftermarket.

Estimates obtained from OLS linear regressions with clustering by subject are presented in Table 2. For both types of auction rules, a null hypothesis that subjects bid no differently when knowing there would be an aftermarket as when there would

[26] Among the possibilities are that a subject might perceive an opportunity to win the auction profitably and then profit further by selling in the aftermarket, which could be perceived as suggesting more aggressive bidding than if there were to be no aftermarket; or that a subject might perceive the aftermarket as a second chance to obtain the asset, which could be perceived as suggesting less aggressive bidding than if there were to be no aftermarket. Both of these possibilities are in equilibrium illusory.

not cannot be rejected at anything vaguely approaching conventional levels of significance.[27]

That subjects did not significantly alter their bidding behavior between auctions known to be followed by an aftermarket and auctions known to be final determiners of payoffs is also suggestive of suitable transparency of the aftermarket structure.

	Table 2	
	Estimates	
	First-Price Data	Second-Price Data
const	-2.346	-1.435
Std. error	(0.567)	(1.629)
Significance	0.001	0.38
β_x	0.045	0.393
Std. error	(0.075)	(0.370)
Significance	0.55	0.29
β_a	-1.242	-5.063
Std. error	(1.431)	(4.738)
Significance	0.39	0.29
# Observations	2215	1075
F test:	1.24	0.58
Significance	0.293	0.562

[27] Failure to reject this null was found in alternatives that did not cluster or added subject fixed effects; alternatives where the bid was the dependent variable and asset value estimate an independent variable were nearly identical.

11. Is Behavioral Efficiency a Distinct Measure?

What can be said about how well behavioral efficiency tracks allocative efficiency? As these experiments used induced values, they can yield insights into the differences between these measures. That is, assume (critically) subjects are all risk-neutral (or identically risk averse), and assume completeness of the induced motivations (in particular, assume away interdependent preferences, nonpecuniary preferences, and satiation in cash). Then in a Pareto-efficient allocation, the asset is acquired by the subject with the highest estimate.

In most observations, when either a first-price or a second-price auction reached a Pareto-efficient allocation, behavioral efficiency was observed in the aftermarket, and the inverse: Pareto-inefficient auctions led to behaviorally inefficient outcomes, mutual gains observed in the aftermarkets.

The two distinctions from tracking were both observed in significant minorities of the observations. [i] In 15% of first-price auctions and 24% of second-price auctions, the efficient acquirer was the high bidder, so the outcome is assumed Pareto-efficient, yet the aftermarket found a mutual gain could arise from a transfer to an outbid rival with a lower estimate of asset value. [ii] For both first-price and second-price auctions, 16% of observations found an inefficient acquirer submitting the high bid and then clicking on an Accept Ask that exceeded all Accept Bids, including that selected by the efficient acquirer whose estimate exceeded his.

While some variant of an endowment effect could lead to the second way in which behavioral efficiency has been found distinct from Pareto efficiency, it bears notice that distinction [i] is completely inconsistent with an endowment effect. More importantly, being able to observe which auction outcomes are Pareto efficient and thus observe these distinctions depends on having induced values and assumed motivational completeness and identical risk tolerances. Using aftermarkets to

observe the size and frequency of shortfalls from behavioral efficiency requires none of these.

12. A Small Field Demonstration

12.1. FIELD CONTEXT: To focus the demonstration on the aftermarket, the initial allocation mechanism is submersed via an assumption that the outcome is production of one unit of public good, with the costs of the first unit's production covered from the experimenter's budget. The aftermarket then considers whether the cost of production of a second unit can be allocated to the perceived mutual benefit of all members of the economy. The aftermarket's construction does not balance the budget, allowing the experimenter to cover a deficit if it is observed that the sum of marginal benefits exceeds production cost of the second unit.

The public good studied is a uniform distribution of small packets of Haribo candy, a product in international distribution and prominent on the shelves of local grocery and convenience stores. It is natural to think of candy as a private good, but in this experiment it was allocated under strict adherence to the definition of a pure public good. That is, [a] there was group exclusion but no individual exclusion in consumption, and [b] there was no rivalry in consumption. Either all subjects in the economy received one unit of candy apiece, or all subjects received two units of candy apiece, depending on whether stated willingnesses-to-pay summed to at least the production cost.

All groups (economies) studied consisted of six subjects. Eighteen subjects (in one session, twelve, due to no-shows) were in the room during a session, so that no subject knew which others were in the same group. Subjects were students recruited by website signup from several campuses of the University of Montpellier. Show-up fees ranged from €3-8, depending on the distance from their home campus to the Experimental Economics Lab at the Richter campus.

Subjects were seated at visually isolated computers. Instructions were passed out and read aloud, questions encouraged and answered.[28] The initial unit of Haribo candy was given to each subject; they were allowed to consume it immediately if they were uncertain of the quality of the candy or for any other reason. They were informed that a second unit would be provided to every member of the group if the sum of the most each group member was willing to pay was at least €1.[29] These were elicited, the second unit provided or not, and subjects paid to the experimenter their cost share for the additional unit (which was necessarily less than the show-up fee).

12.2. METHODOLOGY IMPLEMENTATION: As mentioned, the allocation of 1 unit of public good is treated as if it arose via some allocation mechanism, with the experiment observing an aftermarket. A price clock ticked up on subjects' computer screens, increasing by 2 euro cents, initially every 4 seconds, after 8 euro cents, increasing every 2 seconds. Subjects were asked simply to watch the clock so long as the price was one which they were willing to pay in order to have the group increase public good output from one unit to two, and then to click on the "Accept" button on the screen as soon as the next tick of the clock would yield a price that they were not willing to pay in return for the increase to two units.

Before the clock was run, the outcome function was carefully explained to subjects. If the sum of the six "Accept" prices was at least €1, each subject in the group would be given a second unit of candy, and each subject would pay €1 *minus* the sum of the other five Accept prices (or 0, whichever was larger).

---

[28] An English-language version of the instructions is available at http://harstad.missouri.edu/Instructs/.

[29] Haribo candy was of course available for purchase outside the lab, and a subject's transactions costs of doing so were unknown. Hence, it was important to keep the per-capita threshold for public good production below extra-laboratory prices, so that censoring stated valuations by extra-laboratory availability could not be an issue; cf. Harrison, Harstad and Rutstrom [2004].

This methodology implements the incremental version of the Vickrey-Clarke-Groves mechanism.[30] If a subject is certain of the amount of euros which he would be willing to pay to have the public good output increased from one to two, then it is a dominant strategy to click on the Accept button at the multiple of 2 euro cents nearest his willingness-to-pay.

The incentive compatibility of this methodology warrants the conclusion that the group exhibits a behavioral inefficiency of the initial allocation—of one candy each—if the sum of Accept prices exceeds €1.

12.3. OBSERVATIONS: Thirteen of twenty-three groups exhibited a perceived mutual gain in increasing the distribution of candy from one unit to two units apiece. (This included six of the eleven groups that chose well before lunchtime, and seven of the twelve groups that chose shortly after 1:30 [or three after 3:30 pm], so there is no sign that chronobiology played a role.) The size of behaviorally revealed efficiency gains in these groups ranged from 4% to 72% of production cost (€1), averaging 33%.

The other ten groups found one unit of public good to be behaviorally efficient relative to the sole alternative of two units. The sums of Accept prices in these groups ranged from 56% to 98% of production cost, averaging 86%.

In none of the thirteen groups accomplishing the behavioral efficiency improvement could public good production cost have been covered by a mutually acceptable uniform tax. Rather, only through person-specific pricing could the public good increment be mutually beneficial. Of course, acceptable person-specific price vectors (not uniquely determined in any of the thirteen groups) *could not have been known* to the experimenter, but *were revealed by behavior* in the aftermarket (and the Vickrey-Clarke-Groves selection among those vectors actually used for payment).

---

[30] Named for Vickrey [1961], Clarke [1971] and Groves [1973].

The highest stated willingness-to-pay was €0.54; eighteen of 138 subjects chose an Accept price of €0.04 or less, another forty-five €0.16 or less.

Though there is evidence that it was transparent, it is of course not known whether subjects adopted the dominant strategy of revealing their willingnesses-to-pay. Instructions made it clear that subjects were to evaluate not a second unit of Haribo candy for their own consumption, but a second unit of public good production. Nonetheless, it is unknown whether any subject selfishly placed the same value on second units for all group members as on a private purchase of a second unit. Nor is it known whether any subjects were behaving altruistically, or the extent of any altruistic behavior. It is no more necessary to know their motivations than it would be necessary to know why a consumer purchased a shirt in order to evaluate the allocative efficiency of a shirt market. This aspect justifies treating a laboratory setting as a simple field experiment.

Although the setting was simple almost to the point of contrivance, and the stakes miniscule, this demonstration finds that, at least in the case of public good allocation, the concept can be taken to the field. Whether an adjustment in public good output can be accomplished—via a mutually beneficial decentralization of adjustment costs—can be inferred from observations solely of behavior, provided the aftermarket used to observe those behaviors is appropriately designed. Larger scale, more important field studies can exactly mimic the demonstration offered here, and relate behavioral efficiency observations to the methods used to determine levels of public good output.

12.4. LABORATORY OBSERVATION ON STRATEGIC TRANSPARENCY: Following the demonstration, since the subjects were in an experimental laboratory, a simple induced-value laboratory phase was added. Subjects were given instructions about the allocation of an abstract public good. The only value of this public good was

monetary utility to each individual subject that had been specified by the experimenter.[31]

To mimic the field demonstration, an ad hoc mechanism that was suppressed set initial public good output to 7 units, and each group was asked whether to increase output to 8 units, at an incremental production cost of €3. Each subject was privately told the incremental value v_j to her or him of the increase from 7 to 8 units; the distribution of these incremental values was not announced, although it was announced that the incremental values were not all the same. These six values summed to less than the incremental cost (a random decision, as was the 7-unit starting point); one randomly chosen subject had an incremental valuation equal to €0.7, one equal to €0.06, four equal to €0.42 (that four had the same incremental value was not known to the subjects until results were reported). Division of subjects into groups was via a new random draw, independent of the draw in the field experiment; this feature was announced.

In all other respects, the induced-value procedure was identical to that of the field demonstration: a clock ticked up on all screens (by a multiple of €0.05), subjects were asked to click "Accept" at the highest price they were willing to pay to increase public good output from 7 to 8 units (their incremental value was shown on the screen as the price ticked up), and were told beforehand that an individual group member's personal cost of this increase, which would happen if and only if the sum of Accept prices were at least the production cost, would be the €3 production cost less the sum of the Accept prices of the other five group members. It was carefully explained that their payoff for this decision would be zero if the amount of public good were not changed, and would be their incremental value less the excess of

[31] This assumes selfish preferences, a limitation the field demonstration does not share.

production cost over the sum of the other five Accept prices if public good output were increased.[32]

As before, it is a dominant strategy to set one's Accept price equal to the multiple of €0.05 closest to one's incremental value. Only with incremental values *induced* (thus in the lab, not the field), is it possible to see whether subjects did this. Most did not exactly hit this dominant strategy, although on average the statistic Z = (Accept price – incremental value) was €0.00103, remarkably close to the –€0.015 average it would have been had every subject exactly adopted the dominant strategy.[33] The standard error of Z was nonnegligible, €0.298; the 23 subjects for whom v_j = €0.06 had to click on Accept immediately (at €0.05) were clicking later than this, and averaged Z = €0.38. This was compensated for by the 23 subjects for whom v_j = €0.7, perhaps not waiting for the price clock to reach that high; they averaged Z = –€0.32. Still, over 80% of 138 subjects were within €0.3 of Z = 0, nearly half of those within €0.1. For all twenty-three groups, the six Accept prices summed to less than the €3 production cost, so for all groups, seven units of public good was behaviorally efficient relative to the alternative of eight units. Evidence in the induced-value setting for inexperienced subjects to be *unable* to understand the incentives they faced in the field demonstration is unpersuasive, indeed quite limited.

13. Remarks on Some of the Empirical Issues That May Arise

If there is a limit to the empirical questions that arise with behavioral efficiency, I haven't grasped it. Here's a sampling of those that are clear now.

---

[32] The sessions followed these observations with pilot experiments studying public-good allocation mechanisms that did not bear on the issues of this paper.

[33] Because the Accept bid had to be a multiple of €0.05, in each group the one subject with incremental valuation €0.06 had a dominant strategy Z = –€0.01, and the four subjects with incremental valuation €0.42 had a dominant strategy Z = –€0.02.

Are Pareto-efficient allocations and behaviorally efficient outcomes necessarily distinct? In induced-values settings, under strong assumptions, these can be compared, as in section 11 above.

How large are the magnitudes of shortfalls from behavioral efficiency, and how might these be assessed? In each circumstance where I have been able to envision the outlines of an appropriate aftermarket design, any potential transaction perceived to be mutually beneficial that is observed provides an absolute magnitude of the perceived gain. To put this in percentage terms, as a shortfall from efficiency, requires being able to calculate the Pareto set. In field studies, the best hope is to observe behaviors in the original allocation that indicate the potential size of efficiency gains. Section 9 provides a treatment where such a comparison has a solid theoretical basis, and another with less foundation.

Does knowing that an aftermarket will follow an allocation mechanism affect subjects' behavior in the mechanism? A requirement for aftermarket design is that there is *in theory* no effect (section 4). This will be a potentially important empirical question in every field context, and I expect to have to assess it *de novo*, at least until a large database of field aftermarkets has been compiled. It is possible to design demonstrations and adapt data analysis to shed some light on this, as in section 10. It would be possible in most field settings to surprise subjects with an aftermarket that they almost surely did not anticipate; this will not always be best practice.

Will an aftermarket observe activity just because subjects assume they are supposed to do something? This question arose, for good reason, in reports of experiments observing financial bubbles in labs (Smith, Suchanek and Williams [1988]). The particular "active participation hypothesis" raised by Lei, Noussair and Plott [2001] (that subjects engage in irrational activity because the experimental setup limited rational behavior to inactivity) need not be a concern here, however. In aftermarkets, the anticipated designs will always have subjects do something, in essence engage in

valuation activities. Even if the initial allocation might have been Pareto efficient, there will be a financial incentive to engage in the valuation activities. It should always be possible to structure them so that a zero valuation behavior has nothing to do with a preference for the status quo that was reached in the original allocation mechanism.

Are aftermarket activities mistakes if the initial allocation reached should have been Pareto-efficient? In ordinary cases, it will be difficult if not impossible to label particular aftermarket behaviors mistakes. Only when conducted in induced-values demonstrations will it be possible to determine whether the original allocation reached prior to the aftermarket was in fact Pareto efficient, and even then usually only under assumptions that may well be unverifiable (even the assumptions in section 10 may be insufficient for many situations). In some studies, there could be serious questions about whether the aftermarket design was sufficiently transparent for the subject population, as for example when the subjects are illiterate. When transparency is adequate, I regard it as likely that aftermarket behaviors should be taken at face value.

Might behavioral efficiency determination depend on the structure of the aftermarket used for observation and identification? This will ever be an empirical possibility. When field budgets and subject population sizes permit, multiple aftermarket designs can be tested.

Might endowment effects lead to an inactive aftermarket even though the initial allocation was efficient? This could be imagined, although it is hard to say that inefficiencies exist, let alone identify and quantify them, when an appropriately constructed aftermarket observes inactivity. Plott and Zeiler [2005], [2007] demonstrate that the size of an endowment effect might almost be calibrable. Careful instructions as discussed in the auction experiments above (e.g., avoiding the term "winning bidder" on auction

aftermarkets) yield data (section 11) strongly suggesting that the endowment effect is not a problem in that particular context.[34]

Might other studied psychological biases and behavioral anomalies affect aftermarket observations? To a first approximation, an "anomaly" such as attention to sunk costs, other-regarding preferences, or hyperbolic discounting, may equally impact both an original allocation mechanism and its aftermarket. For many such concerns, there is no reason to believe that they suddenly arise in aftermarkets following an allocation mechanism that went untouched by them. When there is evidence that some particular bias perceived to be relevant to a particular field study can reliably be redressed via education, it may well be best practice to educate first, then run the allocation mechanism followed by the aftermarket. There is considerable evidence suggesting persistence of some biases in the presence of education. Since it is only behaviors that can be observed in field settings, observing aftermarket activity in the presence of such biases may well be the most appropriate way to provide advice for policies that will be promulgated for the population being studied.

14. Field Readiness

A field study of a single-asset auction could exactly mimic the procedures of section 6 to obtain evidence on whether the efficient acquirer won the auction and if not, the size of the inefficiency that arose, even in cases where existence of equilibrium is in doubt (cf. Jackson [2009]) or equilibrium is incalculable (subjects' beliefs about rivals' valuations are unknown) or unobservable (whenever motivations and valuations are unknown to the experimenter). It may be good practice first to familiarize bidders in that field setting with the aftermarket procedures, perhaps by

[34] A referee suggests a broader way in which aftermarket observations might be of limited value in assessing efficiency of outcomes from allocation mechanisms or policy choices: that preferences of subjects may not be stable across an allocation mechanism and the appended aftermarket. If engaging in behaviors that might potentially have transactional impacts *per se* renders preferences unstable, it is unclear that any transactional outcomes can normatively compared (even to autarky), and indeed that any normative economic research—whether theoretical, experimental, or empirical analysis of historical data—can be meaningful.

auctioning off an unrelated, less expensive "demonstration" commodity and then running the aftermarket with the demonstration commodity, prior to conducting the aftermarket to be used to gauge behavioral efficiency.

Fairly straightforward complications of the aftermarket design used here can accommodate observing efficiency shortfalls for mechanisms seeking to allocate multiple homogeneous assets. For example, following a mechanism for allocating two homogeneous assets, potential buyers in an aftermarket could be asked for a pair of Accept Bids if seeking to buy, an Accept ask if one asset won, or a pair of Accept Asks if two, with the rules that whenever an Accept Bid by a rival fell between an Accept Bid and a lower Accept Ask, it set the price for that transaction, and a random price was consulted when necessary. It would not matter whether bidders had single-unit or multi-unit demands. In larger, semi-competitive markets for homogeneous assets, a variant on a call market could serve as an aftermarket (so long as no trader were seeking both to buy more and to sell some of what he had obtained), with the highest quantity where the demand price exceeded the supply price transacted, buyers paying the price of the last accepted supply unit, sellers receiving the (higher) price of the last accepted demand unit, and the experimenter covering the deficit.

Similarly, the aftermarket design used in section 12 is ready for more substantial field usage in any pure-public-good study in which that design is deemed sufficiently transparent for the population to be studied. It is straightforward to observe whether a decrease in public good output from an allocation mechanism outcome Q to $Q - 1$ can remit production cost savings in person-specific, mutually beneficial rebates. For example, as the price clock increases, each subject would be asked to click on the smallest rebate that would compensate for the reduction in output, with VCG rebates implemented if the sum of accepted rebates were at most the production cost savings. Should an increase [decrease] in public-good production be

found mutually advantageous, it would be possible to repeat the aftermarket to see if $Q + 2$ [$Q - 2$] would represent a behavioral efficiency gain relative to $Q + 1$ [$Q - 1$].

15. Concluding Remarks on the Meaning(s) of Behavioral Efficiency

It is a luxury of a parsimonious theory that economists who might disagree about the role of Pareto efficiency—how important is, or perhaps even whether it is desirable—nonetheless agree on the definition of the term and its meaning. I do not see how the terminology of empirical, behavioral studies can have the same luxury. Thus, even if the definition of behavioral efficiency offered here becomes widely accepted, it seems naïve to hope that its meaning will achieve any universal interpretation.

A perhaps less naïve hope is the following. Suppose, starkly, that policy *Y* is observed in a particular context (including a particular subject population and their characteristics) to robustly yield behaviorally efficient outcomes, while alternative policy *Z* in the same context robustly yields outcomes with significantly large shortfalls from behavioral efficiency. Then it might be widely agreed that, at whatever levels of sophistication underlie their perceptions and whatever level of transparency the aftermarket offers, subjects perceive mutual gains from trade that policy *Z* does not capture while perceiving no uncaptured mutual gains from trade following implementation of policy *Y*. Indeed, it might even be widely accepted that advice to policymakers reporting and influenced by this finding could be an improvement over advice reporting and influenced by field experiments that obtain no observations about allocative efficiency. Less starkly, when the observed size of shortfalls from behaviorally efficient outcomes are in context robustly smaller for policy *Y* than for policy *Z*, this might also come to be accepted to play a role in policy advice despite divergent opinions as to its exact meaning.

Appending behavioral efficiency observations can dimensionally enrich field studies.

References

Peter Bohm [1984], Revealing demand for an actual public good, *Journal of Public Economics* 24, 135-51.

Edward Clarke [1971], Multipart pricing of public goods, *Public Choice*, 11, 17-33.

Douglas D. Davis and Charles Holt [1993], *Experimental Economics*, Princeton University Press.

Urs Fischbacher [2007], z-Tree: Zurich toolbox for ready-made economic experiments, *Experimental Economics* 10, 171-78.

David M. Grether, R. Mark Isaac and Charles R. Plott [1979], Alternative methods of allocating airport slots: performance and evaluation, Civil Aeronautics Board Report, Pasadena, Calif.: Polynomics Research Laboratories, Inc.

David M. Grether, R. Mark Isaac and Charles R. Plott [1989], *The Allocation of Scarce Resources: Experimental Economics and the Problem of Allocating Airport Slots*, Westview Press, Boulder.

Theodore Groves [1973], Incentives in teams, *Econometrica* 41, 617-31.

Glenn W. Harrison, Ronald M. Harstad and E. Elisabet Rutstrom [2004], Experimental methods and elicitation of values, *Experimental Economics* 7, 124-40.

Glenn W. Harrison and John A. List [2004], Field experiments, *Journal of Economic Literature* 42, 1009-55.

Ronald M. Harstad [2000], Dominant strategy adoption and bidders' experience with pricing rules, *Experimental Economics*, 3 (2000), 261-80.

Ronald M. Harstad and Robert F. Bordley [1996], Lottery qualification auctions, pp. 157–83 in: M. R. Baye (ed.), *Advances in Applied Micro-Economics,* Volume 6, *Auctions*, JAI Press, Greenwich, Ct.

Matthew O. Jackson [2009], Non-existence of equilibrium in Vickrey, second-price, and English auctions, *Review of Economic Design* 13, 137–45.

John H. Kagel [1995], Auctions: a survey of experimental research, pp. 501–86, in John H. Kagel and Alvin E. Roth, (eds.), *The Handbook of Experimental Economics*, Princeton University Press.

John H. Kagel and Dan Levin [1986], The winner's curse and public information in common-value auctions, *American Economic Review* 76: 894–920.

John H. Kagel, Dan Levin and Ronald M. Harstad [1995], Comparative static effects of number of bidders and public information on behavior in second-price common-value auctions, *International Journal of Game Theory* 24, 297–319.

John H. Kagel and Alvin E. Roth, eds. [1995], *The Handbook of Experimental Economics*, Princeton University Press.

Vivian Lei, Charles N. Noussair, and Charles R. Plott [2001], Nonspeculative bubbles in experimental asset markets: lack of common knowledge of rationality vs. actual irrationality, *Econometrica* 69, 831-59.

Steven Matthews [1977], Information acquisition in competitive bidding processes, working paper, Humanities and Social Science, California Institute of Technology.

Paul R. Milgrom [1981], Rational expectations, information acquisition, and competitive bidding, *Econometrica* 49, 921–43.

Paul R. Milgrom and Robert J. Weber [1982], A theory of auctions and competitive bidding, *Econometrica* 50, 1089–1122.

Roger Myerson and Mark Satterthwaite [1983], Efficient mechanisms for bilateral trading, *Journal of Economic Theory* 29, 265-81.

Charles R. Plott and Vernon L. Smith, eds., [2008], *The Handbook of Experimental Economics Results*, Elsevier North-Holland Publishing.

Charles R. Plott and Kathryn Zeiler [2005], The willingness to pay-willingness to accept gap, the "endowment" effect," subject misconceptions, and experimental procedures for eliciting valuations, *American Economic Review* 95, 530-.

Charles R. Plott and Kathryn Zeiler [2007], Asymmetries in exchange behavior incorrectly interpreted as evidence of endowment effect theory and prospect theory? *American Economic Review* 97, 1449-.

Stephen J. Rassenti, Vernon L. Smith and Robert Bulfin [1982], A combinatorial auction mechanism for airport time slot allocation, *Bell Journal of Economics* 13, 402.

Vernon Smith [1962], An experimental study of competitive market behavior, *Journal of Political Economy*, v. 52.

Vernon Smith [1976], Experimental economics: induced value theory, *American Economic Review*, v. 66.

Vernon Smith, Gerry Suchanek and Arlington Williams [1988], Bubbles, crashes and endogenous expectations in experimental spot asset markets, *Econometrica*, v. 58.

William S. Vickrey [1961], Counterspeculation, auctions and competitive sealed tenders, *Journal of Finance*, 16, 8–37.

Charles Z. Zheng [2002], Optimal auction with resale, *Econometrica* 70, 2197-2224.