Journal of Economic Perspectives—Volume 3, Number 3—Summer 1989—Pages 165–174

Puzzles
The Arbitrage Mirage, Wait Watchers, and More

Barry Nalebuff

This feature begins with three puzzles motivated by the paradox of the "greener envelope," which was described in detail in the Winter 1989 issue. The variants are much simpler and at least one has the flavor of a barroom trick. (I take no responsibility if you are caught.) Answers can be found at the end of the problems. Following these exchange paradoxes are two longer puzzles, also with answers provided at the end. The longer puzzles focus on externality problems: how to design social conventions and toll systems to minimize inefficient waiting in line. The column ends with reader mail, including comments on the welfare optimality of the competitive equilibrium, the pricing policy of the Ocean Club restaurant, Pareto improvements for the U.S. economy, and hints for scoring squash.

Our first T-shirt awards—which are given for the best new puzzles and most innovative answers—go to Morton Davis, Andy Postlewaite, Richard Zeckhauser and Joe Farrell. Well done. For the rest of you, please send your answers, comments, favorite puzzles *and T-shirt size* to me directly: Barry Nalebuff, "Puzzles," Yale School of Organization and Management, Box 1A, New Haven, CT 06520.

Puzzle 1: The Arbitrage Mirage

Morton Davis (The City College of New York) offers the following way to hedge one's bet. "Suppose (for simplicity) it is known that in six months the franc (which is worth exactly one pound today) will either be worth two pounds, or half a pound, and

■ *Barry Nalebuff is Professor of Economics and Management, Yale University, SOM, New Haven, Connecticut.*

that each event has a probability of one-half. A risk neutral Frenchman does well to buy a future in the pound for one franc, since in six months he will be able to convert his pound into either two francs, or half a franc—each with probability one-half: an expected value of 1.25 and an expected profit of 25 percent. An Englishman is in a symmetric position and should buy a future in the franc."

Is this just one of the many gains from trade? Can both make a profit in the long run? Or is this a zero sum game? If so, where does the argument break down (presumably, somewhere in the middle of the English Channel)?

Puzzle 2: Sky Masterson, Watch Out!

Andrew Postlewaite (Univ. of Penn.) relates an old Martin Gardner card trick puzzle. "I open a new deck of cards which you know to be a fair deck. You may then shuffle the deck as much as you'd like. You then gamble on each card turned over in the following manner. You begin with a prespecified amount, say $100. Your first bet will be half of this ($50). If the card turned over is black, you win and get an additional $50; if the card is red, you lose the $50. Whatever amount you are left with, you bet half of what you now have on the turn of the second card, and so on through the deck. Each time you see black results in winning the amount bet, while a red card results in losing the amount bet. Since black and red cards are equally represented, this is a fair bet, correct?"

Puzzle 3: Here's a Second Chance

Morton Davis begins his book "The Art of Decision Making" with a puzzle closely connected to the above card trick. Imagine that the rules are as above, but the win/lose outcome is determined by a coin toss rather than a card draw. With heads you win your bet and with tails you lose. Each round you bet half your assets. After 52 rounds, what is your expected payoff? Are you equally likely to be ahead or behind? What is your conditional expectation given that you are ahead?

If evolutionary forces favor those who earn above average rates of return and work against those earning below average returns, what does this suggest about the evolution of risk-taking tendencies in the population? Compare the survival of two investor types, each with constant relative risk aversion, but one more risk averse than the other.

Puzzle 4: The Thirst to Go First

One of my favorite economic results is due to Refael Hassin (1985) on the optimal way to structure a queue. This puzzle and its solution are based on his paper. However, I must warn you that the puzzle version is almost as long as the original.

Imagine that there is a resource (a drinking fountain) which can service one person at a time. Thirsty individuals arrive randomly at the fountain according to a Poisson distribution.[1] They must decide whether to wait their turn or to leave; if they leave, there is no possibility of coming back later. An individual drinking at the fountain takes a random amount of time, which is again distributed according to a Poisson process.

You may assume that all individuals are identical and that the costs of waiting are linear in time. The question from a social welfare point of view is: how long should the line get? A longer line means more waiting but reduces the chance that the fountain will go unused.[2]

The work of Naor (1969) shows that if the social convention is first-come first-served this will result in an inefficiently large amount of waiting. The reason is that someone who enters the line increases the waiting time for those who come later (although this effect only lasts until the line is empty). A person choosing whether or not to stay in line does not include this externality cost in the calculation of whether or not to wait. The excess incentive to wait in line makes for too long lines and too much standing around.

What alternative social convention results in 100 percent efficient decision-making over whether or not to stay in the queue? An amazing feature of this solution is that without any price mechanism and in a completely decentralized manner, each individual is given just the correct incentive to wait in line from a utilitarian social welfare point of view.

Once you find one answer, you will realize there are many. Is there a "fairest" way to solve the problem? What are the problems with implementing your solution? Can you think of cases when society has adopted one of these efficient rules?

Puzzle 5: Wait Watchers

With the San Francisco–Oakland Bay Bridge in mind, Joe Farrell asked in his Summer 1988 "Puzzles" column how to design traffic tolls that will force drivers to internalize the congestion externality. First, there is the question of calculating the externality a marginal car imposes on other drivers. The answer is remarkably simple: the total externality of an additional car is simply equal to the time remaining before the traffic congestion clears. (For those in doubt, see the solution section.)

Based on this calculation, a toll equal to the externality would simply decline linearly throughout the rush hour(s). Of course, Farrell notes that this calculation

[1]The important feature of the Poisson arrival process is that the distribution of the waiting time remaining before the next person arrives is *independent* of how long we have already waited.

[2]It is simplest to solve the problem when the payoff to being served is 1–0; 1 if you drink for the random time it takes to quench your thirst and zero otherwise. The fit between the fountain story and the mathematics is not perfect. One imagines that drinking is not a memoryless process—at least, for drinking water. Anyway, here you should assume that the expected time the present drinker will remain at the faucet is independent of how long he has been there.

presumes no change in driver behavior and the whole point of the toll is to modify driving patterns. To calculate an optimal toll pattern, it is essential to base the tolls on the externality imposed *in equilibrium*. Thus Richard Zeckhauser offers a follow-up question: with the optimal toll structure in place, do the tolls reflect the waiting time externality imposed by an additional driver?

Answers to Puzzles

Answer to Puzzle 1

The key to resolving this puzzle is to differentiate between the gains from trade and the gains from holding on to one's own currency. And above all else, it is essential to measure gains using some real metric such as loaves of bread or purchasing power rather than a nominal metric such as currency. We will use the loaf of bread as our measure of gains and losses: initially both a pound and a franc buy one loaf of bread.

As is to be expected, the effect of trading francs for pounds is a zero sum game. There are no mutual gains from trade. (Of course, if the players are risk averse, then trade can play an insurance role.) So where is the 25 percent gain coming from?

To solve the problem, we must first specify how it is that the two currencies are expected to change from 1 : 1 into 1 : 2 or 2 : 1 with equal probability. We consider two scenarios in turn. First, look at an asymmetric case. Over the next six months, England expects completely stable prices while France is equally likely to experience either a 100 percent inflation or a 50 percent deflation. Think of the exchange rate as being determined by a foreigner's willingness to pay for a loaf of bread. With a 100 percent inflation in France and stable prices in England, the price of bread will be two francs to one pound. This will be the new exchange rate. Thus, the currencies will trade at either 1 : 2 or 2 : 1 as has been supposed. A Frenchman who buys a future contract on the English pound will find that when he converts his pound back into francs he will be able to buy exactly one loaf of bread. In the event of a 100 percent inflation, his pound translates back into two francs and this is just enough to purchase the loaf of bread at the new price level. And similarly in the event of a deflation. Note that his expected number of francs, 1.25, corresponds exactly to the expected price level six months hence, as this is the average of 2 and 0.5. Thus in real terms, the holder of the stable currency (the pound) makes a zero real return.

The Englishman who buys the future contract on the French franc does in fact end up with an expectation of 1.25 pounds and this buys him 1.25 loaves of bread in England. But note that a Frenchman who does not engage in this future market, but instead holds on to his franc will six months later either be able to buy 2 loaves (with a 50 percent deflation) or one-half a loaf (in the event of a 100 percent inflation). Thus the expected real return to holding on to French currency is 25 percent. That is the source of the Englishman's gain if he buys the future contract.[3]

[3]So why doesn't everyone hold French francs? This expected reward for holding francs must be the market payment for accepting the accompanying risk.

Alternatively, the inflation risk could be completely symmetric across the two countries. Consider a case where exactly one of the two countries will have a 100 percent inflation and the other will have zero percent inflation: which country gets which is equally likely. Holding on to your own currency has a negative real return of 25 percent. Buying the future contract of the other currency leads to an expected 25 percent nominal gain. But the big nominal gain comes when your currency has just suffered an inflation and thus is worth less. The 100 percent nominal gain is exactly offset by the inflation and thus is worth zero percent in real terms. The half chance of a 50 percent nominal loss corresponds to a real loss (since this is the event with no inflation) and this accounts for the expected 25 percent expected real losses. Thus holding on to your currency or trading results in equal expected returns. Once again, there are no expectations of mutual gain from trade.

Answer to Puzzle 2

Wrong. If you fell for this, Postlewaite has a bridge he'd like you to look at in Brooklyn. As he explains, "Not only is this bet not fair, it is not even random. For each red card, your wealth will be multiplied by 0.5 and for every black card it will be multiplied by 1.5. Since multiplication is commutative, the end result will be the original amount of money times $(0.5)^{26}$ $(1.5)^{26}$. Thus the ending amount after one run through the deck is $100 times three-fourths raised to the 26th power. This is approximately five cents (5.64, to be more precise) if I can understand how to use my calculator. The key to seeing the solution is to recognize the difference between sampling with and without replacement."

The solution is perhaps even more transparent in a two card deck. Either you will win and then lose or you will lose and then win. In the first case, you win a small amount ($50) and then lose the bigger gamble of $75; in the latter case, you lose the bigger amount ($50) and then win on the smaller bet of $25. In both cases you end up down $25 with probability one. The dramatic nature of the expected losses suggest that you are unlikely to get through the deck of cards without having the other player suspect he's been had. More subtle and still effective is to proceed only through ten cards and then reshuffle. The expectation is still way in your favor, but now there is at least sufficient randomness so that you may not be discovered immediately.

It is worth noting that there are two elements that are both required to create this paradox. The bet sizes must be positively correlated while the chances of success must be negatively correlated over time. Betting a constant fraction of income, as is the case with constant relative risk aversion, leads to a positive correlation in bet size. The fact that there are exactly 26 winning cards and 26 losing cards forces a negative correlation in performance; because the cards are not replaced, a win today means one less win tomorrow. If the bet sizes were constant or the probabilities independent then it would be a fair bet.

Answer to Puzzle 3

In this case, each gamble is fair so the expected payoff after 52 rounds equals your initial stake: neither a gain nor a loss. However, the chance that you are behind is very close to one. Remember that a pair of a win and loss puts you at $(3/2)(1/2) = (3/4)$ of your initial stake. Thus you have to win almost twice as often as you lose in order to stay even. (Winning twice and losing once places you at 9/8 of your starting amount.) The chance of winning more than 34 times out of 52 is small indeed. Using the normal approximation to the binomial, we have that this chance is about two standard deviations above the mean and thus the chance is roughly 5 percent. But this does not change the fact that the gamble is fair. The 95 percent chance of losses are made up for by the astronomical (or at least exponential) increase in expected value conditional on ending up ahead.

Answer to Puzzle 4

Hassin observes that last-in first-out (LIFO) is an efficient convention for serving the drinking fountain queue. The reasoning is completely intuitive. First off, from a social welfare perspective, it does not matter who waits in line and who gets served. (Here is where we use the assumption that waiting costs are linear with time.) Thus the only question from a social welfare point of view is to ensure that just the right number of people are waiting in line. When the line is an optimal length, society values the waiting of the marginal person *only* if there are no new arrivals before the present person is finished being served. In this case, the line becomes one person shorter, which is one person too short: the presence of the last person is now strictly valued. Thus the marginal person in line should find it just profitable to stay on when he is told that he will be replaced by the next arrival if that person comes before the present person is finished being served.

How can this rule be internalized by a social convention? One way is LIFO: newcomers go to the head of the line.[4] If the last person in line was just indifferent about waiting, then he will choose to leave if any more people arrive before the present server is finished. The reason is that since the new entrant is serviced immediately, the line ahead of him is now one longer and thus he will strictly prefer to leave rather than wait. His calculation about staying versus leaving is the same as the social planner would have him make. In particular, the marginal person's decision about whether or not to stay in line has *no effect* on the waiting times of those who come later. Since there are no externalities, individual cost-benefit calculus comes up with the right social decision.

From a "fairness" point of view, we might want to have the newcomers wait before being served. Hassin observes that any queuing rule where the new entrant goes anywhere but at the end of the line leads to an efficient decision-making. The reason is the same: the marginal person at the end of the line must calculate the benefits of waiting under the conditioning that he will leave if there is a new arrival before the present occupant is finished being served. The standard convention of first-in first-out is the only initial placement rule that is inefficient. Of course, these alternative

[4] If one is worried about interrupting the person being served, it is fine for the new arrival to go second in the queue. This is true even for a one-person queue so that being next in line is also at the end of the queue. See discussion below.

conventions may give rise to some incentive problems as people may want to reenter the queue *from the beginning*; these interesting possibilities and subsequent developments are discussed in Hassin (1985; 1986).

If you think that LIFO is a peculiar way to operate, note that at many stores or ticket offices telephone callers often get priority over people physically waiting in line. Why is it that telephone callers are so often put at the front of the queue? One could pick up the phone and tell them they are the tenth (or whatever) person in line and they will be served after the ninth person waiting has been helped. But that ties up a phone line which might be used for other purposes. Also, the phone caller has no way to monitor his progress, either to update the speed of the line progression or to verify that he will be served when promised. Next time you are annoyed when a phone caller cuts in front of you, remember it is an efficient convention.

Answer to Puzzle 5

One of the bonuses of having guest columnists is that I get to write in with my comments on their puzzles. Here is my trick for calculating the externality caused by another car entering the traffic. Assume that in Farrell's bridge problem the traffic backup starts at 8:00 a.m. and continues to 1:00 p.m. We are asked how much time does one more car add to the sum of everyone else's waiting time. We do not know how long it takes a car to cross the bridge nor any other seemingly essential bits of information. However, when calculating the total waiting time, it does not matter who does the waiting. Thus, imagine that one additional car arrives at 10:30 a.m. Let the driver of this marginal car, instead of proceeding through the toll booth, invisibly stand up his car and let all those behind proceed ahead. In this case, we can assume that no one else is delayed by his presence. Here, it is the marginal driver who bears *all* the extra waiting time. That equals the time it takes for the traffic to clear. If the traffic backup continues until 1:00, the marginal driver arriving at 10:30 a.m. imposes an additional two and one-half hours of waiting time on the population that comes afterwards. This cost is *independent of the distribution of cars that come in the meantime.*

But as some of the readers have been quick to point out (and Joe Farrell notes at the end of his answer) that does not mean one should charge a toll proportional to the externality. Driving patterns will change in the presence of the toll. This brings us to Richard Zeckhauser's question.

With the optimal toll structure in place, *there should be no waiting in line*. Waiting is socially inefficient and should be eliminated by the optimal toll. Thus, if the toll were to reflect the imposed externality measured in waiting time, this would be zero. Of course the toll would then fail to achieve its objective. The externality imposed by a marginal driver is that others must be induced to distort their choice of when (or if) to cross the bridge. This externality simply cannot be measured by waiting times. Instead, times to cross the bridge are sold, like landing slots at airports. The most valuable times are sold for the highest amounts. The optimal toll would be zero before the congestion starts, rise slowly and peak somewhere around the modal optimal time to travel (conditional on no traffic) and then fall as the rush hour passes.

There are several examples which solve for the optimal toll structure, taking into account how drivers shift their schedules in response to the tolls; see Vickrey (1969) and Arnott, Lindsey, and De Palma (1988a, b). Robin Lindsey (Edmonton) has helped convert the early Vickrey paper into a puzzle which you can expect to read about in a future column.

This discussion has focussed on situations where there is a predictable demand to cross the bridge. If the demand includes some random component (say due to the weather or out-of-towners who simply don't know the toll structure) then the solution is more complicated. It doesn't help to have the toll constantly adjusting to the line since this will be too late to influence anyone's behavior. Could it be optimal to have a negative toll before and after the rush hour? While this might encourage some individuals to take the trip when they might not otherwise have done so, it could spread out the rush hour even further thus reducing the probability of a random traffic jam.

Reader Mail

Comments on Puzzles from Summer 1988

Puzzle 5: Are There Pareto Improvements? One of Joe Farrell's challenges was for readers to submit their idea of feasible Pareto improvements. These suggestions continue to be welcome. So far, my favorite suggestion for a Pareto improvement comes from Curt Anderson (Univ. of Minnesota). He suggests we allow left turns on red lights in a manner similar to right turn on red. The idea is to change the convention associated with a traffic light. A person with a green light has absolute right of way. A driver facing a red light may proceed if there are no other cars coming in the orthogonal direction. In theory, no one should lose and we could eliminate some of the wasted waiting at deserted intersections. "Clearly anyone who has ever sat in a long line at a red light with no cross-traffic in sight would benefit from the time savings this interpretation would allow.... Those facing the green light would suffer no loss in their right-of-way and so, would be unharmed by such a change.... Thus, *as long as everyone obeys the law* (i.e., they don't try to proceed on red unless the intersection is truly clear) this change would appear to be a Pareto improvement."

In practice, there is the fear that this might lead to slippage and everyone would treat red lights with the respect that they currently receive in Boston. Perhaps the best comparison is with right turn on red. I look at this as one of the best convention changes in my lifetime. Even so, large metropolitan areas, such as Manhattan, argue that the costs associated with right turn on red outweigh the benefits. The rare opportunity for savings is outweighed by the lower social pressure to stop at red lights. Anderson concludes with a fair question, "Is a policy change which offers a Pareto improvement only if everyone obeys the rules associated with the policy still considered a Pareto improvement?"

Puzzle 1: Competition as a Welfare Minimum. Another of Farrell's puzzles brings interesting mail. He describes the problem of Sylvia, the undergraduate who notices that the competitive equilibrium seems to be a welfare minimum. Social welfare is $W(p) = C(p) + D(p)$. The p that solves $W'(p) = 0$ is the competitive equilibrium as the first-order condition is equivalent to supply equalling demand. But the problem Sylvia noticed is that the second-order conditions are positive, suggesting that the solution is a minimum rather than a maximum.

In his solution, Farrell observed that the right variable is really quantity rather than price. The reason is that the social planner cannot simply set a price and then assume that supply equals demand. At prices other than the competitive equilibrium, supply differs from demand so that one of the two surplus function is incorrect. James Anderson (Boston College) suggests that we can reconcile both positions once we recognize that surplus is based on $x(p) = \min\{\text{Supply}(p), \text{Demand}(p)\}$. It is easy to see that welfare is increasing with x until supply equals demand. No higher x is possible. Thus with the correct rationing rule, welfare is maximized at $x(p^*)$, the competitive equilibrium.

Avinash Dixit writes, "Perhaps your Sylvia just has a wider perspective. For a small country trading with the rest of the world at an exogenous price, p, total surplus is $C(p) + D(p)$, and it is *minimized* when p is such that $D(p) = S(p)$. The interpretation is that of all the worlds our country could find itself trading with, the worst from our viewpoint would be the one which trades at our autarkic price."

Comments on Puzzles from Fall 1987

Puzzle 6: Close Counts. To turn the tables, Joe Farrell writes in with a comment on one of my puzzles. In his continuing quest to improve his squash (which is already excellent) Farrell has discovered a mathematical analysis of squash scoring rules. Alexander, McClements and Simmons (1988) provides some answers to my question about the advantages of scoring according to English versus American rules. In the U.S. squash is played to fifteen, and every point is scored. In the U.K., squash is played to nine points, and points are scored only by the server. The receiver needs to win two consecutive rallies in order to score a point. According to the recent *New Scientist* article, "Suppose that the most probable game scores for three pairs of players under European rules are 9–6, 9–3 and 9–1. A calculation shows that when converted to American rules, the scores would be 15–13, 15–11, and 15–5 respectively. Generally speaking, American rules are far kinder to the weaker player, in the sense that the likelihood of an ignominious defeat is small. Matches would appear much more closely contested, although in fact the probability of a win for either player would not be much altered."

To the extent that both rules lead to the same person winning, rational consumers of squash scores would not care which rule is used. Farrell adds, "It's more a matter of psychology; does it feel good to have a score change after each rally?" Or, does one like scores that sound closer? So Joe, which rule do you want to use when we play?

Comment on Puzzles from Fall 1988

Puzzle 2: Cheaper By the Half Dozen. Finally, we have our latest entry in the "markets are efficient after all" column. This puzzle reported on the Ocean Club restaurant that offered half-portions for less than half price. The proffered explanation was that this seeming irrationality made for good word of mouth (and *JEP*) publicity. Perhaps there was too much of a good thing. Joseph Quinn (Boston College) writes: "You will not be surprised, and perhaps even pleased, to know that the Ocean Club went belly up during the summer, and they closed both of their locations. We all know that coincidence does not imply causation, but there may be a story here nonetheless. The Ocean Club may have been one of those establishments that lost money on every transaction, but made up for it on volume."

■ *This column is based on contributions from JEP readers: Curt Anderson, James Anderson, Morton Davis, Avinash Dixit, Andrew Postlewaite, Joseph Quinn, and Richard Zeckhauser. Behind the scenes, Joseph Farrell and Carl Shapiro help test and improve each puzzle. They have earned my thanks and yours.*

References

Alexander, David, Ken McClements and John Simmons, "Calculating to Win," *New Scientist*, December 1988, 30–33.

Arnott, Richard, Andre de Palma, and Robin Lindsey, "Schedule Delay and Departure Time Decisions with Heterogeneous Consumers," *Transportation Research Record*, 1988a, forthcoming.

Arnott, Richard, Andre de Palma, and Robin Lindsey, "Economics of a Bottleneck," mimeo, University of Alberta, Edmonton, 1988b.

Davis, Morton, *The Art of Decision-Making*. New York: Springer-Verlag, 1985.

Hassin, Refael, "On the Optimality of First-Come Last Served Queues," *Econometrica*, January 1985, *53*, 201–202.

Hassin, Refael, "Consumer Information in Markets with Random Product Quality: The Case of Queues and Balking," *Econometrica*, September 1986, *54*, 1185–1195.

Naor, P., "The Regulation of Queue Size by Levying Tolls," *Econometrica*, January 1969, *37*, 15–24.

Vickrey, William S., "Congestion Theory and Transport Investment," *American Economic Review, Papers and Proceedings*, 1969, *59*, 251–261.

This article has been cited by:

1. . Answers to the chapter questions 197-305. [Crossref]
2. Jeryl L. Mumpower. 2011. Playing Squash Against Ralph Keeney: Should Weaker Players Always Prefer Shorter Games?. *Decision Analysis* **8**:1, 71-77. [Crossref]