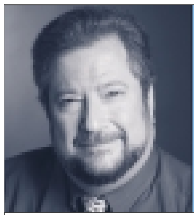
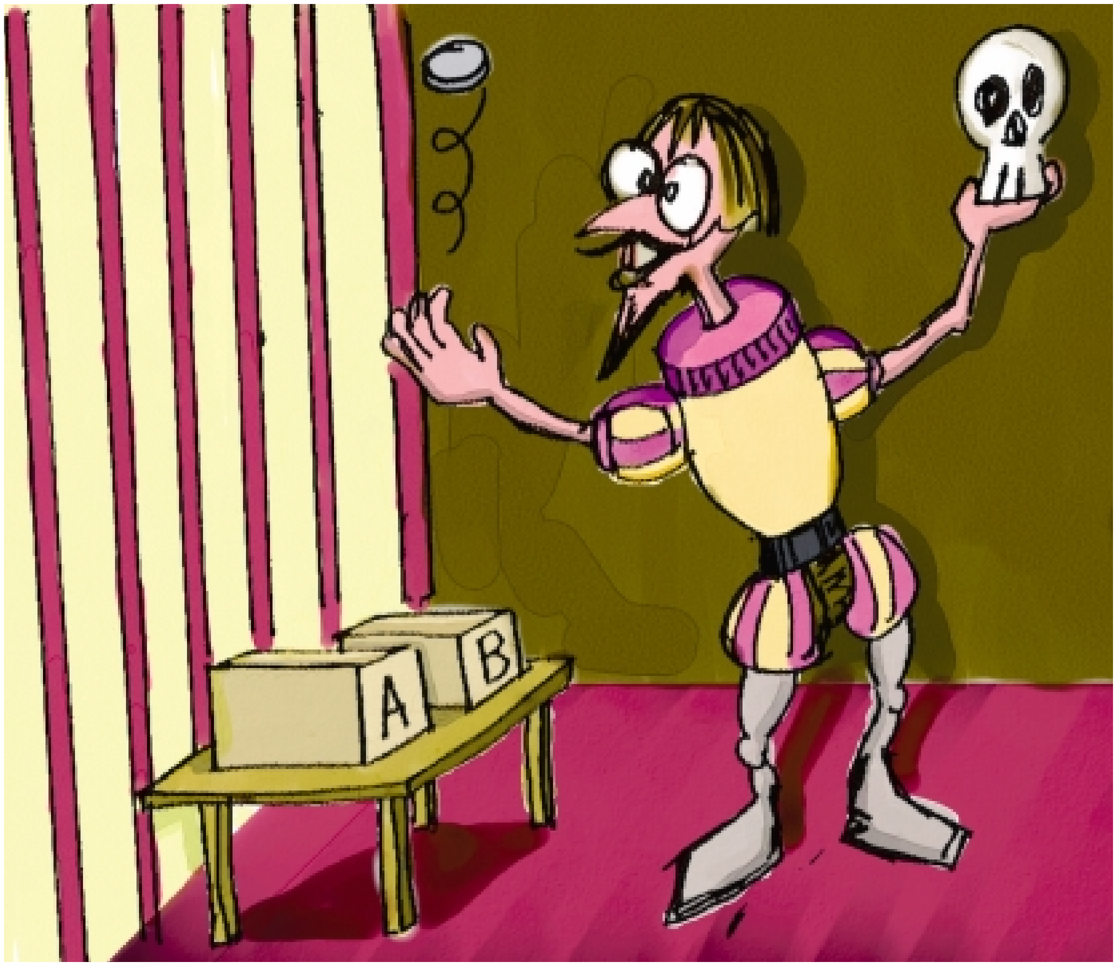


Aaron Brown

Three Puzzles, Three Chapters, Three Gripes

Taking a look at the treatment of Michael Bacharach's last, unfinished work

Michael Bacharach was an important economist, an early champion of some important ideas and an unconventional thinker in a variety of disciplines. But he wasn't William Shakespeare. When he died in 2002, he left an unfinished monograph, *Three Puzzles of Game Theory*, to codify and extend upon his work on how people communicate and cooperate in economic games. This work was too valuable to throw away. It could have been published as-is on the Internet for other researchers to mine as appropriate. It could have been published in book format, with some invited essays to fill in the gaps. Or another researcher could have volunteered to finish it with credit as a co-author.



First gripe: I knew Shakespeare and. . .

Instead, Natalie Gold and Robert Sugden decide to treat it like a newly discovered unfinished Shakespeare play. As editors, rather than co-authors, they make minimal changes to the three chapters Bacharach completed (out of nine planned), adding only annoying parenthetic comments like "Bacharach intended to explain this in Chapter 7, which he never wrote." They preserve an inconsistent and ugly notation that could never have been intended for publication. More likely, Bacharach chose different notations for different problems, some convenient for typewriting and some for handwriting, but would have unified the style and made it more elegant during editing.

The editors open with their own thoughts on what the reader should know before tackling the book, and close with a discussion of how the book might have ended. These are competently done, but do not result in a unified whole. Without rearranging the material or assuming an authoritative voice, it was impossible to do more than stitch together a general survey of the literature and Bacharach's previous work with Bacharach's tightly argued and technical chapters with some loose speculation on future directions. This will satisfy neither casual nor technical readers.

Surprisingly, the result is a great book, *Beyond Individual Choice: Teams and Frames in Game Theory*. It discusses three very simple games that have unsatisfying solutions in classical game theory, and proposes resolutions that go to the heart of finance.

First puzzle: Schelling for focal points

The first puzzle is Schelling games. In 1960, Thomas Schelling asked two players to each choose "heads" or "tails." No communication was allowed. If both players made the same choice, both won $1. If they made different choices, they won nothing. Game theory gives no help here, there's no reason to prefer any strategy to any other. You can't do better than an expected 50 per cent win rate. Yet 87 per cent of players choose "heads," and therefore 77 per cent of players win. This is a robust finding, supported by many subsequent similar studies.

It could be that people just like picking "heads." But that's not it. If you change the payoffs so players win by picking the opposite choice of the other player, they choose "heads" and "tails" close to equally. Obviously people know that other people are more likely to pick "heads," and know other people know that as well.

Bacharach's insight is that this is not just a minor asymmetry when making choices without rational bases. It is an essential survival skill for humans. Humans must cooperate to survive, and full communication is often impossible. If a game is not zero-sum, there is some scope for cooperation, and people are good at exploiting the thinnest edges to communicate. Evolution has allocated large parts of our brains to figuring out, "what would she do if. . .?" and "what would she expect me to do if. . .?"

The book gives no practical examples, but an obvious financial one is technical analysis. Conventional theory emphasizes the zero-sum nature of trading. The average investor earns the average return, any trader who does better must do so at the expense of a trader who does worse. This gives no scope for studying past price patterns to help future decisions in a rational efficient market.

But at any given time, a trader is actually cooperating with roughly half the other traders. Either she wants the price to go up or down, as do many other traders. If humans are good at teasing opportunities for cooperation out of noisy real-world data, it makes sense that traders can read a tape to cooperate with temporary allies.

If you look at things this way, you put a different interpretation on the empirical fact that technical analysis does not work. This is usually cited as evidence for the random walk theory. But the outcome of games can be unpredictable without being random. Chess, for example, has no randomness, but the same player does not win every time. Prices could convey useful signals to different camps of traders, without giving simple money-making rules.

If you believe game theory is a branch of mathematics, you might want to remove Schelling's opportunities for communication by, say, labeling the choices randomly for each player. That way my "heads" may be your "tails." But the point of game theory is making decisions under uncertainty, where the uncertainty involves the actions of an optimizing entity rather than randomness. If you label randomly, you make the actions of the other player random, and this becomes a statistics problem, not a game theory one.

Bacharach ... wants to know not just how people play games, but how they frame real situations into game theoretic form

At the other extreme, if you believe the point of game theory is to predict what people actually do, that is if you think it is a branch of psychology, then there is no puzzle at all. People pick "heads" more than tails and know other people will too. The next job is to compile lists of all kinds of regularities like this.

Bacharach, like me and many other people, thinks game theory is a branch of economics. It should predict not only what people do, but what they should do. A satisfying game theoretic result gives a non-obvious optimum that (a) actually occurs, (b) explains a large amount of observation with a parsimonious theory, (c) is mathematically valid and (d) offers non-obvious, testable, exploitable predictions. In this book, Bacharach goes a step further and demands an evolutionary basis for the mechanism. He wants to know not just how people play games, but how they frame real situations into game theoretic form.

Second gripe: value for money

This is a $35, 200 page book. The production values are poor-to-fair: the text is cramped, the illustrations and backnotes ugly, symbols are used sparingly. The text is well-proofed and copy-edited, however, which is not as common as it should be in technical books.

The 60 pages written by Bacharach are masterfully spare, packed with insight and stripped of nonessentials. While some of the remaining 140 pages are useful, no one would call it spare. The editors cannot mention the concept of unquestioning obedience without quoting at length from *The Charge of the Light Brigade*; the simple idea that there can be more than one perspective requires a drawing of the famous faces/jar gestalt figure (repeated on the title page); a long example of catching a cricket ball, involving nine strategies and eight states of the world, is worked out in excruciating detail, with no apparent relevance (it's the kind of thing an unimaginative instructor might give as an assignment to inject tedium into an exciting subject). Hair-splitting definitions are used when introducing a topic, then never exploited for insight.

It's as if the editors cleared out their desks to make this look like a pop-psychology and a deep mathematical exercise. It is neither. It's 60 pages of path-breaking work in economic game theory, with maybe 20 pages of useful collateral information and 120 pages of none of the above.

Second puzzle: hi lo, hi lo, it's off to work we go

Hi-Lo is similar to the Schelling game, except the payouts for matching are different. Players have two choices. If both pick A, they each get $10, if they both pick B, each gets $1 and if their choices differ they each get nothing. Now the opportunity for cooperation is obvious in the payoffs, you don't need choice labels. And people choose A even more often than they choose "heads," in Schelling games.

>

The puzzle isn't how to play this game. Everyone knows how to do it, and knows that picking A is rational. The puzzle is there is no general proposition in game theory that makes A a better choice than B. People have suggested many rules, but (at least in the eyes of their critics) all either fail to resolve some simple problems, or lead to unacceptable solutions in other problems.

Like Schelling games, it's simple if you reduce it to a probability problem. If neither A nor B is better than the other, the other player will pick randomly, so my expected value from picking A is $5, versus $0.50 from picking B. So I should pick A. If there is some reason to prefer A, it's an even better choice. So I should pick A and the other player should do the same.

But this is not a probability problem. It's easy to see that there are circumstances in which it is rational to choose B. Suppose I know the other player is superstitious and his medium told him today was a good day to avoid "A's" and choose "B's"? Suppose I think he believes this of me?

Here the opportunity for cooperation is even more obvious than in Schelling games. Bacharach believes people solve this problem by reframing it. We notice the opportunity for cooperation. If it is strong enough (which depends not only on the payoffs and labels, but other information) we switch to team-reasoning. We ask "what is the best plan for the team?", then "what is my part in that plan?"

This sounds like a reasonable approximation of how people think when they do things without obvious immediate personal benefit. It also makes some evolutionary sense. People who thought this way formed more successful groups. It may be true that team thinking hinders your success within the group, but this is not obvious. The group may have strategies for punishing or excluding non-cooperators. If not, it may be true that unsuccessful members of successful groups pass on more genetic material, through themselves or kin, than successful members of unsuccessful groups. Finally, we could have a genetically-endowed mechanism that sets up cultural transmission of these traits.

What's not clear is that this explains the hi-lo puzzle. It seems to just redefine the payoffs so each player takes both payoffs into account. That means players are playing a different game than the one specified, rather than the theoretic solution for the specified game is incorrect.

> **We notice the opportunity for cooperation. If it is strong enough ... we switch to team-reasoning. We ask "what is the best plan for the team?", then "what is my part in that plan?"**

Bacharach argues to the contrary, saying the result is not equivalent to a payoff redefinition. For one thing, the original payoffs are what trigger the team-reasoning mechanism. If you kept the sum of the payoffs the same in each outcome state but changed the distribution so that players do not both benefit from the same states, team identification is less likely. Instead, Bacharach claims, there is an agency transformation. The payoffs remain the same, but the players start playing as a team rather than as individuals.

This idea can be taken in the opposite direction. We usually think of people making decisions. But a person is really a series of agents, one at each instant in time. Instant-agent-me might decide to call in sick to work and spend the day skiing, against the interests of both long-term-me and me-as-part-of-the-Morgan-Stanley team, and even me-as-a-part-of-the-honest-humans team (on the other hand, it does support the interests of me-as-a-part-of-the-have-fun-and-don't-take-life-too-seriously team). All of our instant-selves are potential members of lots of teams. Sometimes, as in a mob or dance, many people's instant-selves can fuse and make decisions that none of their long-term-selves and larger teams of long-term selves, would understand. Other times we can make long-term-self-team-decisions that baffle our instant-selves.

Without a theory of agency, game theory doesn't make sense. Bacharach's theory of agency agrees with common sense, but it is not rigorous or well-developed. Among other objections, it clearly demands a parallel theory of communication among agents, and the communication chapter was never written. The main point of this book as it stands is that we need agency theory to explain even the simplest games. Others will have to take up the task of advancing the theory.

Third gripe: critical *p*

There is one important problem in the book which is treated in several places, including two extended examples with numbers. While the Bacharach material treats it only quickly, it seems to be to be clearly correct. The editors, however, go in much more depth, both in the preliminary and concluding material. While they don't commit themselves, they seem to lean toward the wrong answer. Moreover, it seems to be framed as a mathematical question. The mathematics are simple (but not simply explained in the book), the issue is fundamental to Bacharach's game theory.

For a simple example, consider a coordination game in which both players make $1 if they pick the same button, nothing otherwise. There are five buttons arranged in a circle, three red and two blue. One of the buttons is slightly larger than the other four, it is a red button.

One way to frame this problem is "pick a button." In that case there is a 20 per cent chance you will match the other player, whatever she does. Another way is to first pick a color, then a button. Blue is better than red, because if you both pick blue you have a 50 per cent chance of winning, while if you both pick red you have only 33 per cent.

A third framing is pick a size, then pick a button. In this case you pick "large" and if both players do this, win 100 per cent of the time.

One of Bacharach's claims is that people pick the best frame, then the best choice within that frame. They play a metagame that allows them to find a better solution than any game theory based solely on the payoff matrix.

This much is clear. But consider the complication if not everyone notices the size difference. Suppose the probability that a person will notice it is p, and p is known to all noticers. We assume non-noticers will pick a blue button at random.

Clearly, if p is less than 1/3, noticers should pick a blue button as well. That gives them at least a $(1 - p)/2$ chance of winning. Picking the large button wins at most with probability p. If p is greater than 1/2, noticers should pick the large button. If they all do that, they win more than 50 per cent of the time, while 50 per cent is the best you can do picking blue.

The question is what to do for $1/3 < p < 1/2$? Classic game theory is no help. It only tells us that if noticers pick the large button for all $p > 1/3$, we should do this; and if noticers demand a threshold of 1/2, it's rational for us to do this as well.

Mathematically, there is no puzzle at all. If noticers always pick the large button, then a noticer with free will does best by picking the large button when $p > 1/3$. But in this case, the existence of the large button hurts the players. All players would be better off if all buttons were the same size, or, equivalently, if $p = 0$.

Only if $p > 0.5$ does the large button help. So a rational noticer, in Bacharach's theory, will ignore the large button for $p < 0.5$, and he will assume that the other player will ignore it as well. This corresponds with common sense. If there is no rational basis on which to make a coordination decision, we look for some hint to help us coordinate. Once we find one, we don't look for a weaker one. If we can't help but notice a weaker one, we ignore it. If we have five identical buttons arranged in a circle, I might pick the one at 12:00 (if there is one) as the most likely choice. But if one of the buttons is red and flashing "press me," while the others are grey and dull, I'm not worrying about the orientation.

The attractiveness of a coordination frame depends on three things: how likely another player is to choose that frame, how much it narrows the choices down and whether one or more of the indicated choices are likely to be picked for other reasons. Bacharach says that people play games as if they choose frames optimally.

To use 1/3 as the critical value for p, we have to do great violence to the theory. First, we have to take a statistical view rather than a game theoretic one. We rejected that approach at the beginning as not being game theory, because we treat the other player's actions as random rather than optimized. Second, we need the statistical reasoning to be very sloppy. The noticer must ignore every possible frame except color and size. She knows that only p fraction of players notice size, but doesn't consider there may be things she didn't notice that other players will. Finally, she either assumes the other player will act mechanically, picking the large button if it is noticed, or she assumes the other player assumes she will act mechanically.

Third puzzle: stone walls do not a prison make

The most famous puzzle in game theory is the Prisoner's Dilemma. Alice and Bob are caught fleeing an unsuccessful bank robbery. The police do not have physical evidence to convict them of attempted armed robbery, but can manufacture a conviction for resisting arrest if they choose. Prosecutors separate the two and offer them the same deal: confess and go free, while your partner serves ten years. If both confess, each gets nine years (one year off for remorse). If neither confesses, each gets one year for resisting arrest.

Alice can reason that whatever Bob does, she saves one year by confessing. Bob can do the same, so both should confess and serve nine years. Meanwhile, less well-educated crooks are out in one year and admired by each other and the world, while everyone has contempt for Alice and Bob.

Bacharach joins a minority of game theorists who claim it can be rational to refuse to confess. Clearly, in real life it may make sense not to confess. There may be negative consequences to confessing, or positive good feelings to remaining loyal. You may expect to require cooperation in future games. But these argue that the game being played in real life is not Prisoner's Dilemma, not that it is rational to cooperate in the theoretical game. Also, many people (but not all) refuse to confess in experiments in which the explanations above have been blocked.

Bacharach argues that the structure of the game will encourage some people to recognize the opportunity for cooperation, and therefore use team reasoning instead of individual. The best outcome for the team is clearly for both prisoners to refuse to confess, and each individual's part in that is clearly not to confess.

Of course, Bacharach does not say this should happen every time. Each of our instant-selves is a potential member of many teams, we don't know which team (if any) will dominate. The conventional solution assumes both prisoners are fully-cooperating members of their long-term-self teams. But a real prisoner might be as likely to refuse to confess out of momentary bravado, which she will regret later, as loyalty to her partner. She might also be thinking about third parties who will be affected by her choices, or affect her payoff.

This argument becomes more compelling when we think about how socially-determined real payoffs are. Even something as straightforward as money is valued only due to a complex set of social rules. It seems silly to take payoffs and agency-identification as exogenous to the problem. That makes little evolutionary sense, and conflicts with common sense. Bringing them in need not make game theory a branch of psychology, a descriptive field. Bacharach claims convincingly that a rigorous and useful game theory is possible, after agency-identification and cooperation are allowed.

For all my gripes, this is a very important book for anyone interested in quantitative finance. It contains no financial examples, and the work is clearly incomplete (not just because the book itself was unfinished). Bacharach has provided us with 60 pages of rigorous, important thought and the editors, for all their sins, have brought it to us in a useful package. W