Economics and Philosophy, **26** (2010) 27–46
doi:10.1017/S0266267110000040

GAME THEORY: A PRACTITIONER'S APPROACH

Thomas C. Schelling
University of Maryland

To a practitioner in the social sciences, game theory primarily helps to identify *situations* in which interdependent decisions are somehow problematic; *solutions* often require venturing into the social sciences. Game theory is usually about *anticipating* each other's choices; it can also cope with *influencing* other's choices. To a social scientist the great contribution of game theory is probably the payoff matrix, an accounting device comparable to the equals sign in algebra.

In 2005 I received an award 'for having deepened our understanding of conflict and cooperation through game-theoretic analysis'. Does that make me a game theorist? If so, what defines a game theorist?

Notice that 'game theory,' in contrast to almost any other discipline you might think of, has 'theory' in the name of the subject. There are economists, only some of whom are economic theorists; statisticians, only some of whom are statistical theorists; physicists, only some of whom are theoretical physicists; and so on through most disciplines. But game theory has 'theory' in its name. So is a game theorist, like an economist who uses economic theory, someone who *uses* game theory or is a game theorist, like an economic *theorist* who produces economic theory, someone who *produces theory* of the game-theory type?

I am not, or only somewhat, a producer of game-theory theory; I am a user of (elementary) game theory. So I call myself a practitioner, a user, not a creator. (Roger Myerson, in response to the paragraph above, suggested 'game analyst' for people like me.)

This paper is based on the 2008 Witten Lecture in Economics and Philosophy, delivered at Witten–Herdecke University on 27 October 2008.

But that distinction depends on how you want to define game theory, or how the selection committee for my award defined game theory. The two of my publications to which the award committee gave the most emphasis I had published before I knew any game theory. I learned game theory from *Games and Decision*s (1957) by Luce and Raiffa, and have found it – both game theory and the book – immensely helpful for half a century. It gave me techniques to think about, and to articulate, ideas that could benefit from a formal structure.

There are two definitions of game theory. There is a soft one and a hard one. According to the soft one, game theory is the study of how two or more entities – people, governments, organizations – make choices among actions in situations where the outcomes depend on the choices both or all of them make, where each has his or her or its own preferences among the possible outcomes – how they should (might) rationally make their interdependent choices. Each individual needs to anticipate the decisions the others are making. But that means that each needs to anticipate what the others are *anticipating*. And that means anticipating what the others anticipate oneself to be anticipating! This may sound like an infinite regress, but essentially it only means finding a set of expectations that are consistent with each other. Somehow a common expectation of the 'expectable' outcome must be recognized and acted on.

There is another definition, the 'hard' one, that probably reflects, or until recently reflected, the interests of most game theorists, according to which '*Game theory can be defined as the study of mathematical models of conflict and cooperation between intelligent rational decision-makers*' (Myerson, 1991). (My 1975 American Heritage Dictionary, New College Edition, defines game theory as 'the mathematical analysis of abstract models of strategic competition . . .')

The difference is two-fold: the emphasis on 'mathematics' or 'mathematical models', even the exclusivity of mathematics and mathematical models, and the emphasis on 'rational' decision.

There is something ambiguous about the modifier 'mathematical'. I digress here a moment. I consider the greatest invention in the history of mathematics to be the symbol =, the equals sign. The equals sign is an accounting device. If one knows that the two quantities joined by the sign are equal, then one knows that anything added to or subtracted from both sides keeps them equal; any multiplication of both sides by the same quantity, or any raising to a power or extraction of a root leaves them equal; and any manipulation of that kind that leads one side to be equal to zero (conventionally the right side) means that any factor of the other sides is a root of the equation. All algebra – and algebra is embedded in most symbolic mathematics – depends on this simple accounting system.

I understand that the greatest invention in the history of business management is double-entry bookkeeping, double-entry accounting,

according to which all assets and liabilities are related through an equals sign; any discrepancy is identifiable as 'net worth', negative or positive.

The greatest invention in the history of macroeconomics was, until about 75 years ago, the balance of payments accounts; it then, by the 1940s, became the national-income accounts, a quadruple-entry system – combining the double-entry accounts of the two parties to a transaction.

I digressed to introduce what I consider the most helpful invention of game theory *for the social sciences*, the 'payoff matrix'. (Anyone who wants to argue that the concept of 'equilibrium' is the greatest concept will find me willing to argue.) The payoff matrix is usually for two-party situations; it's hard to display three dimensions (although in certain symmetrical situations among many players, a 2×n matrix can often be helpful). It can display multiple choice situations. A two-party two-choice situation may look like this:

	'COLUMN' (chooses *a* or *b*)	
	a	*b*
A	3	1
	3	4
'ROW' (chooses *A* or *B*)		
B	4	2
	1	2

There are four possible outcomes. Each player values the outcomes in the order 4, 3, 2, 1, 4 being most preferred, 1 least preferred. ROW's preferences ('payoffs') are in the lower left corner of each cell: row *B* column *a* is his favourite, row *A* column *b* his least. COLUMN's preferences are in the upper right of each cell. In the abstract logic of game theory, this is a game

https://doi.org/10.1017/S0266267110000040 Published online by Cambridge University Press

if we specify the order of moves. Let it be that choices are simultaneous. We are invited to find the 'solution'.

Note: this 'game' includes only ordinal preferences, that is, the order of preference. It doesn't show whether the worst outcome is much worse than the next to last, compared with the first and second. It also has a somewhat deceptive symmetry, since both players have only the 4, 3, 2, 1 designators of preference. We might have had, for ROW, the numbers 10, 9, 6, 0, and for COLUMN the numbers 10, 3, 2, 1. These cardinal (absolute) numbers would give us more information; whether we need the greater information to 'solve' the game we don't know yet.

Now, my purpose here is not to illustrate how we might try to 'solve' the little game above, but to raise the question, 'Is this mathematics?'

My answer is 'no'. Any possible analysis requires only the ability to tell whether one number is larger than another – not even arithmetic – or in the second case of 10, 9, 6, 0, how much larger – no multiplication, let alone any differential equations! Yet an impressive body of useful game-theory analysis, especially in the social sciences, has been based on simple 2×2 matrices no more complicated than the one above. (Recognizing that 5 is larger than 2 may by definition be mathematics, but not in ordinary parlance.)

A GAME AND SOME CHARACTERISTICS

Actually the one presented above is almost surely the simple matrix most examined in the literature of the social sciences. And it illustrates some useful concepts. One is '*dominance*': one choice is better than the other no matter what the other party chooses. ROW prefers 4 to 3, and 2 to 1, and so may choose row *B* without regard to COLUMN's choice, and similarly for COLUMN's preference for column *b*. Another is '*equilibrium*': if both choose *B* and *b*, neither regrets the choice, neither would unilaterally opt to change it. A third may be called '*payoff dominance*': the payoffs in *A, a* are preferred by both parties to the payoffs they choose in *B, b*; so the outcome they achieve is '*nonoptimal*' but in the absence of some ability to concert they can't get to *A, a*. And that little matrix illustrates the tendency for most situations to involve a combination of conflict and common interest. Their favourite outcomes differ – *B, a* versus *A, b* – but they have a common interest in getting, if possible, *A, a* instead of *B, b*.

Within the confines of abstract game theory, the lower-right cell in the matrix above, yielding the 'dominated' but equilibrium outcome, *B, b*, is usually considered the '*solution*'. Social scientists have devoted a great literature to exploring alternatives to that solution, but they have done so by putting that matrix into some kind of social context, such as communication and contract enforcement, or unilateral promises, or repeated play and reputation, that is, moving into the empirical social sciences and treating that matrix as a kind of core of the situation but not the whole game.

Let me offer another matrix, which will underlie the exposition to come.

		'COLUMN' (chooses *a* or *b*)	
		a	*b*
'ROW' (chooses *A* or *B*)	*A*	2, 2	4, 3
	B	3, 4	1, 1

This one provides neither party a dominant choice; rather the choices are '*contingent*', the best choice for either depends on what the other chooses. And there are *two equilibria*! Lower left and upper right are both outcomes from which neither would unilaterally change. This matrix, like the earlier one, is symmetrical: each faces a choice identical to the other's. If this is a game that has a 'solution', what could the solution be?

A possibility is that, not knowing what COLUMN will do, row should play it safe with the upper row. But if that were a convincing strategy, COLUMN should know it and, expecting row *A*, would choose column *b*, and both would be better off, ROW even better of the two of them. But why not expect COLUMN to be the one to play it safe at *a*? In that case ROW would choose *B*, making both better off, COLUMN more so than ROW. But who should decide to let the other play it safe? If both 'play it safe', they end up at *A*, *a*, second worst for both of them. That doesn't appear a promising basis for a 'solution'.

What I'm going to argue in a moment is that game theory provides a neat way to identify the quantitative characteristics – or, as above, with only the numbers 4, 3, 2, 1, qualitative (ordinal) characteristics – of a situation, but the solution may require going beyond the abstract characterization

https://doi.org/10.1017/S0266267110000040 Published online by Cambridge University Press

in search of more information. What kinds of information? I'm going to propose culture, institutions, precedents, reputations, identifications, even signalling or conversation. And I'm going to use a variation on the above matrix, with cardinal (absolute) values, rather than ordinal, as below:

	'COLUMN' (chooses *a* or *b*)	
	a	*b*
'ROW' (chooses *A* or *B*)		
A	8, 8	10, 9
B	9, 10	1, 1

This matrix is symmetrical: each faces identical choices and outcomes. One outcome is very bad. Two are equilibria: they are clearly the best outcomes, but which should be chosen?

(In some situations there are many, even infinitely many, equilibria, sometimes with identical payoffs; the problem then, if choices must be taken independently without communication, is one of pure coordination, to identify some hint or signal or suggestion or precedent or 'rule' that both parties, or all parties, recognize as a common expectation. An example: I've arranged for all the students in my class to be admitted to special seating at the inaugural parade in Washington. They will be recognized because, as I told them, they will all give the same password. But I neglected to give them the password! I've told the person in charge that my students can be easily identified because they will all give the same password. My students have no way of communicating with each other; they must all give the same password to be admitted. How do they all choose the same password? Classroom experience finds most of them able to 'solve' this problem.)

https://doi.org/10.1017/S0266267110000040 Published online by Cambridge University Press

A non-equilibrium is at upper left. Evidently there is a basis for playing it safe, to avoid the lower right. But if either could be expected to play it safe, or for any other reason to choose *A* or *a*, the other should make both better off by choosing *B* or, as the case may be, *b*. If each expects the other to play it safe, they choose B and b, and get a very bad outcome.

A SITUATION, EXEMPLIFIED BY THE 'GAME'

Now I want to give an 'interpretation' of that matrix, a very common situation that all of us are acquainted with, have encountered numerous times, and see what a 'solution' – either an expectable outcome or a somehow preferred outcome – might depend on.

Here's the interpretation. Two cars approach an intersection at right angles to each other, one to the other's left, one to the other's right, going at similar speeds and evidently going to arrive at the intersection simultaneously at those speeds. If both continue without slowing, they will arrive at the same time and some sort of collision, or at least a jolting stop, is inevitable. If both slow down, both lose time, and they still haven't 'solved' the problem. (We then have another 'game' to play: both stopped, waiting for the other.) If either one slows down while the other proceeds at speed, they get a 'solution' in the sense that there is no superior alternative, each gets at least his second best, but one loses a little time relative to the other.

Each needs to anticipate whether the other will slow down or continue at the same speed. And their expectations have to be consistent: if each expects the other to slow down, they will collide (or come to a jolting stop); if each expects the other to continue at speed, they lose time unnecessarily. If one slows down and the other continues, they have 'cooperated'. A favourable outcome is necessarily asymmetrical: in the abstract we don't know which of the two, out of generosity, modesty or caution will slow down or how the other may come to believe he or she can proceed with confidence.

The answer is not in the matrix. The *question* is nicely formulated in the matrix, the *answer* is not.

So where is the answer? Or where do we look for it? Where do the two drivers look for it? How likely are they to find *the* solution, or *a* solution?

There are some interesting possibilities. One is that something in the situation points to an obvious pair of choices. We can call that pair of choices a 'solution'. For example, maybe if one were going too fast to stop in time and the other could see that only he could avert collision, the latter would know that only he could 'solve' the problem. But I proposed they were going at the same speed.

Maybe there is a 'clue' that both can recognize: one's type of vehicle or one's style of driving indicates recklessness and the other chooses caution. Maybe, at their speeds, the danger is not bodily injury but only damage to the car, and one car is new and expensive and the other is old and already

out of shape and has little to lose in collision at moderate speed. Maybe it is visible that one car carries children and the other only the driver, and the latter knows that the car with children will cautiously slow down.

Another possibility is that there is a rule, some convention, known to both drivers and known to be known to both, that indicates who is to slow down. 'Ladies first' may be a possibility, if gender is visible. That the car to the other's right has the privilege may be a known rule. Of course, red and green lights can provide a rule known to all drivers. Some of these rules may need enforcement; some may depend on courtesy and good citizenship. And some may be self enforcing. The red and green lights make it dangerous to claim an intersection when the light is red: the other car will be expecting clear passage. (The coloured lights probably need no legal status; all that's required is the discriminating signal.) Similarly, the rule that the car to the other's right goes first may be so well known that it is dangerous to contest it. If it is widely believed that taxis are willing to risk moderate damage to the vehicle in order to complete the journey quickly, and especially if it is known that taxi drivers believe in the universality of that belief, taxis will take the right of way and others will acknowledge it by slowing down.

Of course, 'ladies first', or 'taxis first', or 'new car slows down' is an asymmetric solution, a discriminatory solution, but even the one discriminated against benefits from there being a recognizable solution.

I had an experience in Beijing a quarter-century ago that dramatized the self-enforcement concept. Bicycles were swarming on a wide avenue and I tried to cross, watching the oncoming bicycles with a view to navigating safely among them. At one point it became clear to me that I'd better halt briefly. The result was 14 cyclists tumbling to the pavement. I was later told the 'rule'. Keep moving at constant speed in a straight line and pay no attention to the bicycles; they will be counting on your steady movement in a straight line, and any departure will only confuse them and cause the kind of multiple collisions I had innocently provoked.

Then there is the possibility of signalling. A nice asymmetrical possibility is putting one's hand out the window and gesturing to the other to proceed: only the driver on the other's right can do that, the other driver's left hand would not be visible to the driver on the right [in the USA].

An interesting question is whether cooperation could be more effective if drivers approaching the intersection could speak to each other. The technology surely exists, though it's not generally available. My suspicion is that with communication the bargaining may not prove efficient. Both parties may be demanding; both may declare unilaterally, both may be so generous that both offer to slow down like two people waiting for each other to be the first through a door. Too many options may make it harder to concert on one.

One car – we hope it's not both – might have an indicator, perhaps on the windshield or on the license plate, indicating 'I always slow down' or

'I always demand right-of-way'. Is this a credible declaration? I believe it must be: if the car says it slows down, the driver certainly doesn't want the other to slow down, and if the car claims right-of-way, it certainly doesn't intend to slow down. Deception doesn't lead to solution.

(We might add another option: speed up. This would be effective if the other driver either slowed down or maintained speed, but the enhanced risk if both speed up might seem to make it too dangerous to be worth considering. Still, if it is so dangerous that no sensible driver would consider it, maybe that makes it a safe option! Game theory suggests that if it is so risky it has to be considered. It is somewhat like Yogi Berra's remark, 'Nobody goes there any more, it's too crowded'.)

I hope this little exploration of a familiar common situation of simultaneous choice among two participants will have displayed the kinds of situations that game theory explores. I hope it also suggests, as I have experienced as a 'game analyst', that game theory is great at exploring situations, less able to provide solutions in the abstract. I usually find that the 'solution' to one of these reciprocal-choice problems depends on the details, not simply on the abstract model. The model – the abstract definition of the situation in terms of the payoffs – can be absolute; the solution, if one is found, is usually contingent on who or what the parties are and what they know about each other, culture and institutions, history and precedent, and what is common knowledge.

It is fair to ask why these contextual matters cannot be treated as part of game theory. Fifty years ago I wrote about commitments of various kinds, using matrices, and had the presumption to subtitle my article, 'Prospectus for a reorientation of game theory'(Schelling, 1958). I had no discernible influence.

I hope also that it may be evident why I consider the payoff matrix to be such a useful product of game theory, perhaps more evident as I proceed. It is usually limited to two-party situations – it's hard to deal with matrices in more than two dimensions. It is most often used, in the social sciences, with two-choice situations, as above, but it can accommodate many-choice situations. But that is beyond my present purpose.

ANTICIPATE VS. INFLUENCE

In the work for which I received the award, my interest has been less in problems of reciprocal anticipation, like the one we just worked, than in how the parties may attempt to influence each other's behaviour, the choices each other makes. (Even in our little traffic example, we saw the possibility of signalling.) This subject arises in behaviour among nations, in industrial disputes, in criminal law, in bargaining over a purchase, in encouraging and disciplining children or pets, in extortion and blackmail, and even, as we saw, in negotiating automobile (or bicycle) traffic.

An important way of influencing another's behaviour, I observed, was by influencing the other's expectation of one's own behaviour. And that, I observed, was often accomplished by determining one's own behaviour in advance, in a manner visible to the other, or communicable to the other in a credible way. And one could attempt to determine one's own behaviour either unconditionally or conditional on the other's response. I was especially impressed with the role of *commitment*, of becoming *committed* to a course of action: 'I'm going through' in our traffic example, or 'I'm slowing down'.

In the traffic example, 'I'm going through' was believable because it wouldn't serve the driver's purpose unless it were the driver's evident intention. And it was an unconditional commitment, not one contingent on the other's slowing down.

But 'come one step closer and I'll shoot' may require convincing that the gun is real, that the gun is loaded, and that one actually would dare to fire at the target. And it is conditional: it implies 'and if you don't, I won't'. It is what I call a 'threat'. In calling it a threat I mean that what is threatened is what one would prefer not to do. If one would actually prefer to shoot in that contingency – that the intruder keep coming – it would be a warning, a statement revealing a truth that the intruder surely wants to know. (Bluffing, of course, is always a possibility; one can fire a shot to prove the gun is loaded, but the willingness to shoot is not so easily proven.)

Another commitment is the *promise*. A promise may be conditional or unconditional. 'If you clean your room I'll take you to the ball game', or 'I'll be home in time to take you to the ball game'. The latter will assure that the child will be home.

What is interesting is how difficult it may be to take a firm unconditional commitment, to issue a believable threat, or to make a believable promise. I'll illustrate some of these points using one of the matrices above, but to illustrate the importance, and the possible difficulty, of making appropriate believable promises, let me describe a television show by Alfred Hitchcock.

An old man passes a jeweller's store in the darkness of early morning just as three men emerge with bags in their hands. They've evidently robbed the store. Just as they are about to get into their getaway car the leader turns back to the old man and says, 'Sorry, old man, but we can't afford to leave any witnesses alive'. The old man says he wouldn't tell anything; the robber says that the old man could identify them if they were ever caught, and they can't afford that possibility. The old man says, 'Wait, there must be some alternative. Does anyone have a knife?' When someone produces a scissors he says, 'Put my eyes out'. One way to make a promise believable is to make it impossible to renege – if you can find a way!

To show how promises may afford a 'solution,' let's look at that original matrix above, the one that led to an inferior outcome. Here it is again:

		'COLUMN'	
		(chooses *a* or *b*)	
		a	*b*
		3	4
'ROW' (chooses *A* or *B*)	*A*	3	1
		1	2
	B	4	2

It doesn't matter who has to choose first, or who gets to choose first, or whether they choose simultaneously; the outcome is ineluctably at the lower right, with payoffs of 2 and 2, *as long as all they care about is the numerical payoffs*. But suppose ROW can make a believable promise, and so can COLUMN. Here's what ROW can propose (and, to make it simple, let's assume that all he has to do is to say 'I promise'). He says, 'I promise that if you promise to choose *a* I shall choose *A*'. Note that this is a conditional promise: he doesn't promise to choose *A*, he promises to choose *A if* COLUMN makes a corresponding (unconditional) promise to choose *a* before ROW makes his choice. Both have to be capable of believable promises. But if they have or can arrange that capability they can 'solve' the problem presented by the matrix.

My interest is mainly in how and when and under what circumstances some individuals – people, governments, corporations, unions, political parties – can actually make believable promises. But it is important to see, as in the above matrix, what a great difference it can make.

Before moving to the subject that most interests me, when or how or who can actually make such promises, let me suggest a way to display the promise in the little matrix. One way to express that promise is this: ROW says, 'I promise that if you promise to reduce your payoff in *A*, *b* from 4 to 2, I promise then to reduce my payoff in *B*, *a* from 4 to 2'. If COLUMN

https://doi.org/10.1017/S0266267110000040 Published online by Cambridge University Press

so promises, and ROW keeps his promise, and so does COLUMN, the matrix becomes:

		'COLUMN' (chooses *a* or *b*)	
		a	*b*
'ROW' (chooses *A* or *B*)	*A*	3, 3	1, 2
	B	2, 1	2, 2

Here neither has a 'dominant' choice; but there are two equilibria, one obviously preferred by both. In effect, each incurred a 'penalty' on violating his promise, the reduction from 4 to 2 in the event of defecting on the promise. And both gain from their abilities to make believable promises.

There are many 2 × 2 matrices that can demonstrate the payoff structures that make promises helpful, to the one promising or to both, or that make threats helpful to the threatener, or that display when a promise, or a threat, cannot work alone, but a combination of threat and promise can help the one making the threat and the promise. (For example, a blackmail threat needs the promise not to reveal if the victim pays; the nuclear deterrent threat needs to assure that no attack is forthcoming except in retaliation.) But what we need now is to examine what determines who can make a believable promise, under what circumstances, using what facilities or what institutions may be available.

This is where we depart from the payoff structure and engage in empirical study. Some might say this is where we depart from game theory and move into social science. It certainly ceases to be mathematical, if it ever was.

https://doi.org/10.1017/S0266267110000040 Published online by Cambridge University Press

For an extended discussion of who can make believable promises, threats, or unconditional commitments, in what circumstances and within what institutions, I must refer you to a book of mine, *The Strategy of Conflict* (Schelling, 1960). Here I'll just mention a sample of circumstances. Begin with the promise.

In the movie, *The Princess Bride*, the reluctant maiden is wed to the evil prince in a bumbling ceremony that is interrupted by an attack on the castle. In the confusion she meets the hero, whom she loves, and confesses that all hope is lost, she is married. The hero demands, 'Did you say 'I do''? After some reflection she is pretty sure that that part of the ceremony got omitted in the battle for the castle. 'Then you are not married. You can't be if you didn't say 'I do'.. 'I do' in a marriage ceremony is part of a formula – a 'performative utterance' in the terminology of Austin (1962) and a 'speech act' in the terminology of Searle (1969) – that changes the legal status of the woman's relation to the man who must also say 'I do' (and not merely 'yes').

But even a child, told he or she will receive a specific reward for good behaviour, is likely to say, 'Promise?'. Uttering the word 'promise' affects one's relation to the child; merely offering the reward is somewhat less serious than is promising. And failure to keep the promise makes future attempts at promise less rewarding.

Signing, with witnesses, a legal contract is a method of reciprocal promising. Having a reputation for keeping promises is an asset; issuing a promise stakes that reputation on the fulfilment. Being known to believe in a deity that enforces promises provides one a capacity to invoke penalty on defections; 'cross my heart and hope to die' or 'may God strike me dead' can be credible. Offering a tangible pledge, as a forfeit, may work; one offers one's guitar to the pawnshop to guarantee repayment of the loan. In earlier times, hostages were offered, or exchanged.

There are occasionally 'mechanisms' for arranging commitments. In the 1930s many national labour unions in the USA with numerous locals that might find themselves engaged in a strike arranged 'strike insurance', according to which any local union engaged in a strike could count on financial contributions from all the other locals, to help avert the worst consequences of lost wages. The intention, I understand, was originally only to share the burdens among the locals. But the effect was to make striking so much less costly to the striking union that its threat to persist in the strike became much more credible than if there had been no financial recourse. The bargaining position – the 'commitment' to persevere – was thus enhanced.

In 1950 President Truman proposed that the Congress authorize the stationing of seven army divisions in Germany, to bolster NATO's defence. The question arose, could seven added divisions make enough difference to a possible successful defence against a Soviet-bloc invasion? Secretary of State Dean Acheson, questioned by the US Senate, explained that what the

seven divisions could do was not to make possible an effective defence of Western Europe; that was not feasible for the time being. What they could do was to guarantee that if 300 000 American young men were killed or captured, the war could not stop there; it would escalate ineluctably to a higher level of warfare. They were the commitment, the pledge, the hostages.

The subjects of 'commitment', of 'threat' and of 'promise' I've elaborated at length in my books and will not go into here to any further length. Game theorists have, at least until recently, been reluctant to deal empirically with the social mechanisms that make commitments possible, possibly because they prefer the abstract beauty of the mathematical domain. I have in mind the legal arrangements; the cultural restraints and demands; religious beliefs and practices; tribal, ethnic, neighbourhood and kinship relations; organizations' rules; the possibilities for secrecy or revelation; communication and technological facilities and restraints; even intrafamily relations.

Even the most elementary game theory, as exemplified by the 2×2 matrix, can help to elucidate the scope of commitment, promise, or threat. Consider this matrix:

		'COLUMN' (chooses a or b)	
		a	b
'ROW' (chooses A or B)	A	2, 1	1, 2
	B	0, 0	0, 0

With that matrix suppose that COLUMN chooses first, ROW's choice to follow, but ROW gets to commit himself before COLUMN chooses. (How he manages to commit himself, assuming he can, is itself a big subject.) If ROW does nothing, COLUMN chooses a, ROW then chooses A (because 1

https://doi.org/10.1017/S0266267110000040 Published online by Cambridge University Press

is better than 0), and COLUMN gets 2. If ROW can threaten, 'If you choose *a* I shall choose *B*', COLUMN faces zero if he chooses *a*, so he chooses *b*, and ROW's payoff is 2 rather than 1. Of course, ROW has to commit himself to that threat in a manner believable to COLUMN and communicate the threat. The matrix is independent of however it is that ROW can make the threat; but matrices as simple as this one can illustrate what payoffs make a promise effective, or a threat, or an unconditional commitment to a choice.

One more illustration:

		'COLUMN' (chooses *a* or *b*)	
		a	*b*
'ROW' (chooses *A* or *B*)	*A*	2	4
		5	4
	B	1	5
		1	0

Here again the game is that COLUMN has first choice, with ROW's choice to follow, but ROW can make an advance commitment. Note that if ROW does nothing in advance, COLUMN chooses *a* knowing that ROW will then choose *A*, to get 2 rather than 1, COLUMN getting 5. If ROW commits himself unconditionally to *B*, COLUMN chooses *a*, and both do badly. If ROW Threatens '*B* if *a*' to force COLUMN to choose *b*, it won't work because COLUMN will get 0 if he chooses *b*, row *B* providing ROW 5 vs. 4. ROW has to promise, 'If *b*, not *B*' while threatening 'If *a*, then *B*'. Neither the threat nor the promise alone will induce COLUMN to choose *b*, yielding ROW a score of 4 rather than the 2 he would achieve in the absence of the coupled promise and threat. So, '*B* if *a*, *A* if *b*' is the effective commitment. Again, how the commitment is arranged is outside the domain of game theory, narrowly defined.

https://doi.org/10.1017/S0266267110000040 Published online by Cambridge University Press

SITUATIONS

The logic of choice is central to game theory, but what game theory is often critical for – again I'm thinking about the invention of matrices – is identifying situations in which choice is somehow puzzling, problematic or challenging, and in identifying how many distinct situations of a certain kind there may be.

In the 1970s I participated in a two-week seminar on arms control in Aspen, Colorado. My contribution was to examine 'the different motives that can lead two countries to bargain about armaments'. I wanted to classify the alternative preferences about possession or non-possession of weapons that arms bargainers could have and to see what kinds of bargains were compatible with different preferences, what understandings and misunderstandings were likely, what the role of ignorance or deception might be, what bargains might need some kind of enforcement.

I ignored, for simplicity, differences within governments and imputed to each of two governments a strict preference order. I considered only binary choices – to have or not to have the particular weapon (e.g. ABM, biological weapons, weapons in space) and excluded consideration of capabilities that each might prefer the other to have (such as secure control over weapons, or false-alarm-free warning systems) and considered only weapons that each preferred the other *not* to have. This simplified formulation allowed all possibilities to be contained in a 2 × 2 matrix:

	'COLUMN' (chooses *have* or *not*)	
	have	*not*
'ROW' (chooses *HAVE* OR *NOT*)		
HAVE	*a* : *A*	*c* : *B*
NOT	*b* : *C*	*d* : *D*

https://doi.org/10.1017/S0266267110000040 Published online by Cambridge University Press

The payoffs to ROW are the capital letters, those to COLUMN the lower case. We can think of the letters as representing four ordinal numbers, say 4 (best), 3 (next best), 2 (third best) and 1 (least favoured). The only restriction on how we place the four numbers, 4, 3, 2 and 1, for each of the two parties, is that: $B > A$, $D > C$, $b > a$ and $d > c$, i.e. that whether one side has it or not it prefers the other not to. With that restriction we have six configurations of payoffs for ROW, six for COLUMN, and 36 possible matrices. They correspond to:

1. We want it *whether or not you have it*, and prefer *both* have to *neither* have it;
2. " and prefer *neither* have to *both* have it;
3. We want it *if and only if* you *have* it;
4. " you *do not have* it;
5. We do *not* want it whether or not you have it, but prefer, *if one* has it, that *we* do.
6. " " " *you* '.

I then eliminated 4 and combined 5 and 6. I could dream up a weapon to which 4 might apply, but it is far-fetched; and with 4 eliminated, the difference between 5 and 6 was of no consequence. I gave them names: 2 were called 'YES' and 1 'YES!' with exclamation point, 3 is IFF ('if, but only if') and 5 and 6 consolidated were simply 'NO'. (Note that YES! means we want it so badly we wouldn't trade it away, YES without the exclamation mark means we want it whether or not you have it but would prefer we both abstain – probably what people most often have in mind for arms control.)

That left four configurations for each party, or 16 combinations. For example, the matrix for the combination IFF/YES:

		'They' YES	
		have	*not*
	HAVE	2 , 3	1 , 2
'We' IFF	*NOT*	4 , 1	3 , 4

https://doi.org/10.1017/S0266267110000040 Published online by Cambridge University Press

'They' have a dominant choice: they prefer to have it – whatever 'it' is – whether we have it or not, what I called the YES configuration. 'We' have a contingent choice: we want it if they have it, not if they don't, the IFF configuration. There is one equilibrium, the upper left cell, which is fairly unsatisfactory for both of us. But they know that if they choose *not*, we shall choose *NOT*. That is, we follow them, and they'd prefer *neither* to *both*.

If we were actually not IFF but NO, and they knew it, they'd choose *have*. But if we could deceive them into believing we were IFF, we'd both choose *NOT* (respectively *not*) and, paradoxically we'd both be better off! (Or if we were NO but could arrange, perhaps via legislation, a commitment to IFF, we could achieve the lower right outcome.)

Of the 16 matrices, four of them are symmetrical, 12 (like the one above) asymmetrical. You can easily construct them for yourself, or find them in Schelling (1984).

I then considered each of the 16, whether any bargain could be in the joint interest, whether a bargain had to be enforceable, whether a bargain was unnecessary, whether it made a difference if a second weapon were brought into consideration (making either a trade, or a coercive choice, possible), what misunderstandings could arise to make a bargain unachievable, what commitments might be incurred (e.g. by legislation), how 'bargaining chips' might be useful or mischievous, whether deception could play a role. All these issues are not 'game theory', by traditional definition; but game theory provides the matrices with which to begin the analysis.

I've gone into some detail in order to demonstrate what I mean by game theory having to do with the analysis of *situations*, and how matrices can be essential to any exhaustive exploration of those situations.

An ironic footnote here is that the essay I produced was being published by a journal, which shall be nameless, the editor of which insisted, 'Tom, you've got to delete all the matrices'. I said nobody could follow the argument without the matrices; he said he'd rather have his subscribers perplexed than intimidated. Out they went, but I shortly published the comprehensible version elsewhere.

RATIONALITY

A final point about the role of 'rationality', that I mentioned in introducing the 'hard' definition of game theory. I said that game theory involved two or more decision makers with independent preferences among the outcomes, and said the entities could be persons, governments, organizations etc. There is a substantial literature on whether a committee, or a corporation, or a government, or even a team, can be expected to display the 'rationality' of a single rational 'individual'. Some of that literature is

quite abstract and theoretical, involving, for example, the majority-vote paradox and the Arrow 'impossibility' theorem for collective choice. Some is empirical, as in the work of Allison (1971 [1999]) on the Cuban missile crisis.

If game theory is confined to 'rational choice' strictly defined, it cannot claim to apply to collective decision except in special cases. Studies of the Cuban Missile Crisis, of the Executive Committee of the Kennedy Administration, make clear that on matters of extreme strategic importance, the United States Government is not a 'unitary' individual but a mixture of cabinet officers, military chiefs, senators and congresspersons, as well as a president. I don't believe it is wise to define game theory in a way that excludes the most important decisions that nations may engage in.

Irrationality, or I should say 'irrationalities', plural, can be manageable in game theory as long as the nature of the particular 'irrationality' can be identified. For example, does one participant not understand the other's language; is one deaf; does one suffer from claustrophobia or some other debilitating phobia; is one a small child, or an elderly person suffering dementia; is one known to be susceptible to overwhelming rage; is one known to be subject to a particular superstition; does one suffer a form of amnesia; is one addicted to a substance; is one innocent of any statistical sophistication, incapable of thinking probabilistically; is one for the time being inebriated or under the influence of a sedative or other drug? Or, of course, both of them. And are either the 'irrational' individual, or the other party, or both, aware of the particular 'irrationality' and how it affects decisions? Camerer (2003) explores many ways that idiosyncratic behaviour can be accommodated in game theory.

Not long ago I underwent a minor medical procedure that entailed a mild anaesthesia. I was not allowed to drive myself home. I was instructed not to operate any machinery for the rest of the day. And I was especially instructed, in writing, not to sign any documents! I think game theory should be able to handle my temporary 'irrationality'.

A truncated form of game theory, involving threats, promises and coordinating signals, can even be applied to the training of animals, who clearly are not fully 'rational' by human standards.

REFERENCES

Allison, G. T. 1971. *Essence of Decision: Explaining the Cuban Missile Crisis*. 2nd edn, with Philip Zelikow, 1999. Boston: Little, Brown & Co.

Austin, J. L. 1962. *How to Do Things with Words*. London: Oxford University Press.

Camerer, C. F. 2003. *Behavioral Game Theory*. Princeton, NJ: Russell Sage Foundation, Princeton University Press.

Luce, R. D. and H. Raiffa 1957. *Games and Decisions*. New York: John Wiley and Sons.
Myerson, R. B. 1991. *Game Theory: Analysis of Conflict*. Cambridge, MA: Harvard University Press.
Schelling, T. C. 1958. The strategy of conflict: prospectus for a reorientation of game theory. *Journal of Conflict Resolution* II, No. 3.
Schelling, T. C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
Schelling, T. C. 1984. *Choice and Consequence*. Cambridge, MA: Harvard University Press.
Searle, J. R. 1969. *Speech Acts*. Cambridge: Cambridge University Press.