War and peace

Robert J. Aumann*

Center for the Study of Rationality, and Department of Mathematics, The Hebrew University, 91904 Jerusalem, Israel

"Wars and other conflicts are among the main sources of human misery." Thus begins the *Advanced Information* announcement of the 2005 Bank of Sweden Prize in Economic Sciences in Memory of Alfred Nobel, awarded for Game Theory Analysis of Conflict and Cooperation. So, it is appropriate to devote this lecture to one of the most pressing and profound issues that confront humanity: that of war and peace.

I would like to suggest that we should perhaps change direction in our efforts to bring about world peace. Up to now, all the effort has been put into resolving specific conflicts: India–Pakistan, North–South Ireland, various African wars, Balkan wars, Russia–Chechnya, Israel–Arab, etc., etc. I'd like to suggest that we should shift emphasis and study war in general.

Let me make a comparison. There are two approaches to cancer. One is clinical. You have, say, breast cancer. What should you do? Surgery? Radiation? Chemotherapy? Which chemotherapy? How much radiation? Do you cut out the lymph nodes? The answers are based on clinical tests, simply on what works best. You treat each case on its own, using your best information. And your aim is to cure the disease, or to ameliorate it, in the specific patient before you.

And, there is another approach. You don't do surgery, you don't do radiation, you don't do chemotherapy, you don't look at statistics, you don't look at the patient at all. You just try to understand what happens in a cancerous cell. Does it have anything to do with the DNA? What happens? What is the process like? *Don't* try to cure it. Just try to *understand* it. You work with mice, not people. You try to make them sick, not cure them.

War has been with us ever since the dawn of civilization. Nothing has been more constant in history than war. It's a phenomenon, it's not a series of isolated events. The efforts to resolve specific conflicts are certainly laudable, and sometimes they really bear fruit. But, there's also another way of going about it—studying war as a general phenomenon, studying its general, defining characteristics, what the common denominators are, what the differences are. Historically, sociologically, psychologically, and—yes—*rationally*. Why does *homo economicus*—rational man—go to war?

What do I mean by "rationality"? It is this:

*A person's behavior is **rational** if it is in **his** best interests, given **his** information.*

With this definition, can war be rational? Unfortunately, the answer is yes; it can be. In one of the greatest speeches of all time—his second inaugural—Abraham Lincoln said: "Both parties deprecated war; but one would make war rather than let the nation survive; and the other would accept war rather than let it perish. And the war came."

It is a big mistake to say that war is irrational. We take all the ills of the world—wars, strikes, racial discrimination—and dismiss them by calling them irrational. They are not necessarily irrational. Though it hurts, they may be rational. If war is rational, once we understand that it is, we can at least somehow address the problem. If we simply dismiss it as irrational, we can't address the problem.

Many years ago, I was present at a meeting of students at Yale University. Jim Tobin, who later was awarded the Nobel Memorial Prize, was also there. The discussion was freewheeling, and one question that came up was: Can one sum up economics in one word? Tobin's answer was "yes"; the word is *incentives*. Economics is all about incentives.

So, what I'd like to do is an economic analysis of war. Now this does *not* mean what it sounds like. I'm not talking about how to finance a war, or how to rebuild after a war, or anything like that. I'm talking about the *incentives* that lead to war, and about building incentives that prevent war.

Let me give an example. Economics teaches us that things are not always as they appear. For example, suppose you want to raise revenue from taxes. To do that, obviously you should raise the tax rates, right? No, wrong. You might want to *lower* the tax rates. To give people an incentive to work, or to reduce avoidance and evasion of taxes, or to heat up the economy, or whatever. That's just one example; there are thousands like it. An economy is a game: the incentives of the players interact in complex ways, and lead to surprising, often counterintuitive results. As it turns out, the economy really works that way.

So now, let's get back to war, and how *homo economicus*—rational man—fits into the picture. An example, in the spirit of the previous item, is this. You want to prevent war. To do that, obviously you should disarm, lower the level of armaments. Right? No, wrong. You might want to do the exact opposite. In the long years of the cold war between the U.S. and the Soviet Union, what prevented "hot" war was that bombers carrying nuclear weapons were in the air 24 hours a day, 365 days a year. Disarming would have led to war.

The bottom line is—again—that we should start studying war, from all viewpoints, for its own sake. Try to understand what makes it happen. Pure, basic science. *That* may lead, eventually, to peace. The piecemeal, case-based approach has not worked too well up to now.

Now I would like to get to some of my own basic contributions, some of those that were cited by the Prize Committee. Specifically, let's discuss repeated games, and how they relate to war, and to other conflicts, like strikes, and indeed to all interactive situations.

Repeated games model long-term interaction. The theory of repeated games is able to account for phenomena such as altruism, cooperation, trust, loyalty, revenge, threats (self-destructive or otherwise)—phenomena that may at first seem irrational—in terms of the "selfish" utility-maximizing paradigm of game theory and neoclassical economics.

That it "accounts" for such phenomena does not mean that people deliberately choose to take revenge, or to act generously, out of consciously self-serving, rational motives. Rather, over the millennia, people have evolved norms of behavior that are by and large successful, indeed optimal. Such evolution may actually be biological, genetic. Or, it may be "memetic"; this word derives from the word "meme," a term

Author contributions: R.J.A. wrote the paper.

*E-mail: raumann@math.huji.ac.il.

Editor's Note: This article is a version of Robert Aumann's Nobel Lecture, "War and Peace." The 2005 Nobel Prize in Economics was awarded jointly to Robert Aumann and Thomas Schelling for having enhanced our understanding of conflict and cooperation through game theory analysis. The Nobel Foundation has graciously granted us permission to print this article to help broaden its exposure.

coined by the biologist Richard Dawkins to parallel the term "gene," but to express social, rather than biological, heredity and evolution.

One of the great discoveries of game theory came in the early 1970s, when the biologists John Maynard Smith and George Price realized that strategic equilibrium in games and population equilibrium in the living world are defined by the same equations. Evolution—be it genetic or memetic—leads to strategic equilibrium. So what we are saying is that in *repeated* games, strategic equilibrium expresses phenomena such as altruism, cooperation, trust, loyalty, revenge, threats, and so on. Let us see how that works out.

What do I mean by "strategic equilibrium"? Very roughly, the players in a game are said to be in *strategic equilibrium* (or simply *equilibrium*) when their play is *mutually optimal*: when the actions and plans of each player are rational in the given strategic environment—i.e., when each knows the actions and plans of the others.

For formulating and developing the concept of strategic equilibrium, John Nash was awarded the 1994 Nobel Memorial Prize in Economics, on the 50th anniversary of the publication of John von Neumann and Oskar Morgenstern's *Theory of Games and Economic Behavior* (1). Sharing that prize were John Harsanyi, for formulating and developing the concept of *Bayesian* equilibrium, i.e., strategic equilibrium in games of incomplete information; and Reinhard Selten, for formulating and developing the concept of *perfect* equilibrium, a refinement of Nash's concept, on which we will say more below. Along with the concepts of *correlated* equilibrium (2, 3), and *strong* equilibrium (4), both of which were cited in the 2005 Prize announcement, the above three fundamental concepts constitute the theoretical cornerstones of noncooperative game theory.

Subsequent to the 1994 prize, two Nobel Memorial Prizes were awarded for *applications* of these fundamental concepts. The first was in 1996, when William Vickrey was awarded the prize posthumously for his work on auctions. (Vickrey died between the time of the Prize announcement and that of the ceremony.) The design of auctions and of bidding strategies are among the prime practical applications of game theory; a good—though somewhat dated—survey is ref. 5.

The second came this year—2005. Professor Schelling will, of course, speak and write for himself. As for your humble servant, he received the prize for applying the fundamental equilibrium concepts mentioned above to *repeated* games. That is, suppose you are playing the same game G, with the same players, year after year. One can look at this situation as a single big game—the so-called *supergame* of G, denoted G^∞—whose rules are, "play G every year." The idea is to apply the above equilibrium concepts to the supergame G^∞, rather than to the one-shot game G, and to see what one gets.

The theory of repeated games that emerges from this process is extremely rich and deep; good—though somewhat dated—surveys are refs. 6–8. In the few minutes that are available to me, I can barely scratch its surface. Let me nevertheless try. I will briefly discuss just one aspect: the *cooperative*. Very roughly, the conclusion is that

Repetition Enables Cooperation.

Let us flesh this out a little. We use the term *cooperative* to describe any possible outcome of a game, as long as no player can *guarantee* a better outcome for himself. It is important to emphasize that in general, a cooperative outcome is *not* in equilibrium; it's the result of an agreement. For example, in the well-known "prisoner's dilemma" game, the outcome in which neither prisoner confesses is a cooperative outcome; it is in neither player's best interests, though it is better for both than the unique equilibrium.

An even simpler example is the following game H: There are two players, Rowena and Colin. Rowena must decide whether both she and Colin will receive the same amount—namely 10—or whether she will receive 10 times more, and Colin will receive 10 times less. Simultaneously, Colin must decide whether or not to take a punitive action, which will harm both Rowena and himself; if he does so, the division is cancelled, and instead, each player gets nothing. The game matrix is

	Acquiesce	**P**unish
Divide **E**venly	10, 10	0, 0
Divide **G**reedily	100, 1	0, 0

The outcome (**E**,**A**), yielding 10 to each player, is a cooperative outcome, as no player can guarantee more for himself; but as in the prisoner's dilemma, it is not achievable in equilibrium.

Why are cooperative outcomes interesting, even though they are not achievable in equilibrium? The reason is that they are achievable by contract—by agreement—in those contexts in which *contracts are enforceable*. And there are many such contexts. For example, a national context, with a court system. The Talmud (Avot 3,2) says,

הוי מתפלל בשלומה של מלכות,
שאלמלא מוראה,
איש את רעהו חיים בלעו.

"Pray for the welfare of government, for without its authority, man would swallow man alive." If contracts are enforceable, Rowena and Colin can achieve the cooperative outcome (**E**,**A**) by agreement; if not, (**E**,**A**) is for practical purposes unachievable.

The cooperative theory of games that has grown from these considerations predates the work of Nash by about a decade (1). It is very rich and fruitful, and in my opinion, has yielded *the* central insights of game theory. However, we will not discuss these insights here; they are for another Nobel Memorial Prize, in the future.

What I do wish to discuss here is the relation of cooperative game theory to repeated games. The fundamental insight is that repetition is like an enforcement mechanism, which enables the emergence of cooperative outcomes *in equilibrium*—when everybody is acting in his own best interests.

Intuitively, this is well known and understood. People are much more cooperative in a long-term relationship. They know that there is a tomorrow, that inappropriate behavior will be punished in the future. A businessman who cheats his customers may make a short-term profit, but he will not stay in business long.

Let's illustrate this with the game H. If the game is played just once, then Rowena is clearly better off by dividing **G**reedily, and Colin by **A**cquiescing. (Indeed, these strategies are *dominant*.) Colin will not like this very much—he is getting almost nothing—but there is not much that he can do about it. Technically, the *only* equilibrium is (**G**,**A**).

But in the supergame H^∞, there *is* something that Colin can do. He can *threaten* to **P**unish Rowena for ever afterwards if she ever divides **G**reedily. So it will not be worthwhile for her to divide greedily. Indeed, in H^∞ this is actually an equilibrium in the sense of Nash. Rowena's strategy is "play **E** forever"; Colin's strategy is "play **A** as long as Rowena plays **E**; if she ever plays **G**, play **P** forever afterwards."

Let's be quite clear about this. What is maintaining the equilibrium in these games is the *threat of punishment*. If you like, call it "MAD"—mutually assured destruction, the motto of the cold war.

One caveat is necessary to make this work. The discount rate must not be too high. If it is anything over 10%—if $1 in a year is worth less than 90 cents today—then cooperation is impossible, because it's still worthwhile for Rowena to be greedy. The reason is that even if Colin punishes her—and himself!—for ever afterwards, then when evaluated today, the entire eternal punishment is worth less than $90, which is what Rowena gains today by dividing greedily rather than evenly.

I don't mean just the monetary discount rate, what you get in the bank. I mean the personal, subjective discount rate. For repetition to engender cooperation, the players must not be too eager for immediate results. The present, the now, must not be too important. If you want peace now, you may well never get peace. But if you have time—if you can wait—that changes the whole picture; *then* you may get peace now. It's one of those paradoxical, upside-down insights of game theory, and indeed of much of science. Just a week or two ago, I learned that global warming may cause a cooling of Europe, because it may cause a change in the direction of the Gulf Stream. Warming may bring about cooling. Wanting peace now may cause you never to get it—not now, and not in the future. But if you can wait, maybe you will get it now.

The reason is as above: The strategies that achieve cooperation in an equilibrium of the supergame involve punishments in subsequent stages if cooperation is not forthcoming in the current stage. If the discount rates are too high, then the players are more interested in the present than in the future, and a one-time coup now may more than make up for losses in the sequel. This vitiates the threat to punish in future stages.

To summarize: In the supergame H^∞ of the game H, the cooperative outcome (**E**,**A**) is achievable in equilibrium. This is a special case of a much more general principle, known as the *Folk Theorem*, which says that *any* cooperative outcome of *any* game G is achievable as a strategic equilibrium outcome of its supergame G^∞—even if that outcome is not an equilibrium outcome of G. Conversely, every strategic equilibrium outcome of G^∞ is a cooperative outcome of G. In brief, for any game G, we have

THE FOLK THEOREM: **The cooperative outcomes of G coincide with the equilibrium outcomes of its supergame G^∞.**

Differently put, repetition acts as an enforcement mechanism: It makes cooperation achievable when it is not achievable in the one-shot game. Of course, the above caveat continues to apply: In order for this to work, the discount rates of all agents must be low; they must not be too interested in the present as compared with the future.

There is another point to be made, and it again relates back to the 1994 prize. John Nash got the prize for his development of equilibrium. Reinhard Selten got the prize for his development of *perfect* equilibrium. Perfect equilibrium means, roughly, that the punishment is *credible*; that *if* you have to go to a punishment, then after you punish, you are still in equilibrium—you do not have an incentive to deviate.

That certainly is *not* the case for the equilibrium we have described in the supergame H^∞ of the game H. If Rowena plays **G** despite Colin's threat, then it is *not* in Colin's best interest to punish forever. That raises the question: In the repeated game, can (**E**,**A**) be maintained not only in strategic equilibrium, but also in *perfect* equilibrium?

The answer is yes. In 1976, Lloyd Shapley—whom I consider to be the greatest game theorist of all time—and I proved what is known as the *Perfect Folk Theorem*; a similar result was established by Ariel Rubinstein, independently and simultaneously. Both results were published only much later (9, 10). The Perfect Folk Theorem says that in the supergame G^∞ of any game G, any cooperative outcome of G is achievable as a *perfect* equilibrium outcome of G^∞—again, even if that outcome is not an equilibrium outcome of G. The converse of course also holds. In brief, for any game G, we have

THE PERFECT FOLK THEOREM: **The cooperative outcomes of G coincide with the perfect equilibrium outcomes of its supergame G^∞.**

So again, repetition acts as an enforcement mechanism: It makes cooperation achievable when it is not achievable in the one-shot game, even when one replaces strategic equilibrium as the criterion for achievability by the more stringent criterion of *perfect* equilibrium. Again, the caveat about discount rates applies: In order for this to work, the discount rates of all agents must be low; they must not be too interested in the present as compared with the future.

The proof of the Perfect Folk Theorem is quite interesting, and I will illustrate it very sketchily in the game H, for the cooperative outcome (**E**,**A**). In the first instance, the equilibrium directs playing (**E**,**A**) all the time. If Rowena deviates by dividing **G**reedily, then Colin punishes her—plays **P**. He does not, however, do this forever, but only until Rowena's deviation becomes unprofitable. This in itself is still not enough, though; there must be something that motivates Colin to carry out the punishment. And here comes the central idea of the proof: If Colin does not punish Rowena, then Rowena must punish Colin—by playing **G**—for not punishing Rowena. Moreover, the process continues—any player who does not carry out a prescribed punishment is punished by the other player for not doing so.

Much of society is held together by this kind of reasoning. If you are stopped by a policeman for speeding, you do not offer him a bribe, because you are afraid that he will turn you in for offering a bribe. But why should he not accept the bribe? Because he is afraid that you will turn him in for accepting it. But why would you turn him in? Because if you don't, he might turn you in for not turning him in. And so on.

This brings us to our last item. Cooperative game theory consists not only of delineating all the possible cooperative outcomes, but also of choosing among them. There are various ways of doing this, but perhaps best known is the notion of *core*, developed by Lloyd Shapley in the early fifties of the last century. An outcome x of a game is said to be in its "core" if no set S of players can *improve* upon it—i.e., assure to each player in S an outcome that is better for him than what he gets at x. Inter alia, the concept of core plays a central role in applications of game theory to economics; specifically, the core outcomes of an economy with many individually insignificant agents are the same as the competitive (a.k.a. Walrasian) outcomes—those defined by a system of prices for which the supply of each good matches its demand (e.g., refs. 11 and 12). Another prominent application of the core is to *matching* markets (e.g., refs. 13 and 14). The core also has many other applications; for surveys, see refs. 15–20.

Here again, there is a strong connection with equilibrium in repeated games. When the players in a game are in (strategic) equilibrium, it is not worthwhile for any one of them to deviate to a different strategy. A *strong* equilibrium is defined similarly, except that there it is not worthwhile for any *set* of players to deviate—at least one of the deviating players will not gain from the deviation. We then have the following

THEOREM (4): The core outcomes of G coincide with the strong equilibrium outcomes of its supergame G^{∞}.

In his 1950 thesis, where he developed the notion of strategic equilibrium for which he got the Nobel Memorial Prize in 1994, John Nash also proposed what has come to be called the *Nash Program*—expressing the notions of cooperative game theory in terms of some appropriately defined noncooperative game; building a bridge between cooperative and noncooperative game theory. The three theorems presented above show that repetition constitutes precisely such a bridge—it is a realization of the Nash Program.

We end with a passage from the prophet Isaiah (ch. 2, vss. 2–4):

וְהָיָה בְּאַחֲרִית הַיָּמִים, נָכוֹן יִהְיֶה הַר בֵּית
יְי בְּרֹאשׁ הֶהָרִים, וְנִשָּׂא מִגְּבָעוֹת, וְנָהֲרוּ
אֵלָיו כָּל הַגּוֹיִם. וְהָלְכוּ עַמִּים רַבִּים וְאָמְרוּ,
לְכוּ וְנַעֲלֶה אֶל הַר יְי, אֶל בֵּית אֱלֹהֵי
יַעֲקֹב,וְיֹרֵנוּ מִדְּרָכָיו, וְנֵלְכָה בְּאֹרְחֹתָיו ; כִּי
מִצִּיּוֹן תֵּצֵא תוֹרָה, וּדְבַר יְי מִירוּשָׁלָם.
וְשָׁפַט בֵּין הַגּוֹיִם, וְהוֹכִיחַ לְעַמִּים רַבִּים ;
וְכִתְּתוּ חַרְבוֹתָם לְאִתִּים, וַחֲנִיתוֹתֵיהֶם
לְמַזְמֵרוֹת ; לֹא יִשָּׂא גוֹי אֶל גּוֹי חֶרֶב, וְלֹא
יִלְמְדוּ עוֹד מִלְחָמָה.

"And it shall come to pass that- . . . many people shall go and say, . . . let us go up to the mountain of the Lord. . . . And He will teach us of His ways, and we will walk in His paths. . . . And He shall judge among the nations, and shall rebuke many people; and they shall beat their swords into ploughshares, and their spears into pruning hooks; nation shall not lift up sword against nation, neither shall they learn war any more."

Isaiah is saying that the nations can beat their swords into ploughshares when there is a central government—a Lord, recognized by all. In the absence of that, one *can* perhaps have peace—no nation lifting up its sword against another. But the swords must continue to be there—they cannot be beaten into ploughshares—and the nations must continue to *learn* war, in order *not* to fight!

I am grateful to Prof. Nicolaus Tideman for pointing out an error in a previous version.

1. von Neumann J, Morgenstern O (1944) *Theory of Games and Economic Behavior* (Princeton Univ Press, Princeton).
2. Aumann RJ (1974) *J Math Econ* 1:67–96.
3. Aumann RJ (1987) *Econometrica* 55:1–18.
4. Aumann RJ (1959) *Ann Math Study* 40:287–324.
5. Wilson R (1992) in *Handbook of Game Theory, with economic applications*, eds Aumann RJ, Hart S (Elsevier, Amsterdam), Vol 1, pp 305–329.
6. Sorin S (1992) in *Handbook of Game Theory, with economic applications*, eds Aumann RJ, Hart S (Elsevier, Amsterdam), Vol 1, pp 71–107.
7. Zamir S (1992) in *Handbook of Game Theory, with economic applications*, eds Aumann RJ, Hart S (Elsevier, Amsterdam), Vol 1, pp 109–154.
8. Forges F (1992) in *Handbook of Game Theory, with economic applications*, eds Aumann RJ, Hart S (Elsevier, Amsterdam), Vol 1, pp 155–177.
9. Aumann RJ, Shapley LS (1994) in *Essays in Game Theory in Honor of Michael Maschler*, ed Megiddo N (Springer, New York), pp 1–15.
10. Rubinstein A (1994) in *Essays in Game Theory in Honor of Michael Maschler*, ed Megiddo N (Springer, New York), pp 17–28.
11. Debreu G, Scarf H (1963) *Int Econ Rev* 4:235–246.
12. Aumann RJ (1964) *Econometrica* 32:39–50.
13. Gale D, Shapley LS (1962) *Am Math Mon* 69:9–15.
14. Roth A, Sotomayor M (1990) *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis* (Cambridge Univ Press, Cambridge, UK).
15. Anderson RM (1992) in *Handbook of Game Theory, with economic applications*, eds Aumann RJ, Hart S (Elsevier, Amsterdam), Vol 1, pp 413–457.
16. Gabszewicz JJ, Shitovitz B (1992) in *Handbook of Game Theory, with economic applications*, eds Aumann RJ, Hart S (Elsevier, Amsterdam), Vol 1, pp 459–483.
17. Kannai Y (1992) in *Handbook of Game Theory, with economic applications*, eds Aumann RJ, Hart S (Elsevier, Amsterdam), Vol 1, pp 355–395.
18. Kurz M (1994) in *Handbook of Game Theory, with economic applications*, eds Aumann RJ, Hart S (Elsevier, Amsterdam), Vol 2, pp 1153–1192.
19. Peleg B (1992) in *Handbook of Game Theory, with economic applications*, eds Aumann RJ, Hart S (Elsevier, Amsterdam), Vol 1, pp 397–412.
20. Young HP (1994) in *Handbook of Game Theory, with economic applications*, eds Aumann RJ, Hart S (Elsevier, Amsterdam), Vol 2, pp 1193–1236.