Stochastic games

Eilon Solan[a,1] and **Nicolas Vieille**[b,1]

[a]Department of Statistics and Operations Research, School of Mathematical Sciences, Tel Aviv University, Tel Aviv 6997800, Israel; and [b]Department of Economics and Decision Sciences, HEC Paris, 78 351 Jouy-en-Josas, France

In 1953, Lloyd Shapley contributed his paper "Stochastic games" to PNAS. In this paper, he defined the model of stochastic games, which were the first general dynamic model of a game to be defined, and proved that it admits a stationary equilibrium. In this Perspective, we summarize the historical context and the impact of Shapley's contribution.

game theory | stochastic games

The 1950s were the decade in which game theory was shaped. After von Neumann and Morgenstern's *Theory of Games and Economic Behavior* (1) was published in 1944, a group of young and bright researchers started working on game theory, each of whom published papers that opened new areas. In 1950, John Nash published two papers, one on the concept of Nash equilibrium (2) and the other on the bargaining problem (3). In 1953, Lloyd Shapley published two papers, one on stochastic games (4) and the other on the Shapley value for coalitional games (5). In 1957, Harold Kuhn (6) provided a formulation for extensive-form games, and Duncan Luce and Howard Raiffa published their book *Games and Decisions: Introduction and Critical Survey* (7). These researchers and others, like Kenneth Arrow, Richard Bellman, and David Blackwell, interacted, worked and played together, exchanged ideas, and developed the theory of dynamic games as we know it today.

While in graduate school at Princeton, Shapley befriended Martin Shubik and John Nash. The former became a well-known economist, and the latter defined the concept of Nash equilibrium, the most prevalent solution concept in game theory and economic theory to date. In 2013, Shapley was awarded the Nobel Prize in economics, 8 y after Nash was awarded the prize.

1. Stochastic Games

Stochastic games model dynamic interactions in which the environment changes in response to players' behavior. In Shapley's words, "In a stochastic game the play proceeds by steps from position to position, according to transition probabilities controlled jointly by the two players" (4).

A stochastic game is played by a set of players. In each stage of the game, the play is in a given state (or position, in Shapley's language), taken from a set of states, and every player chooses an action from a set of available actions. The collection of actions that the players choose, together with the current state, determine the stage payoff that each player receives, as well as a probability distribution according to which the new state that the play will visit.

Stochastic games extend the model of strategic-form games, which is attributable to von Neumann (8), to dynamic situations in which the environment changes in response to the players' choices. They also extend the model of Markov decision problems, which was developed by several researchers at the RAND Corporation in 1949–1952, to competitive situations with more than one decision maker.

The complexity of stochastic games stems from the fact that the choices made by the players have two, sometimes contradictory, effects. First, together with the current state, the players' actions determine the immediate payoff that each player receives. Second, the current state and the players' actions influence the choice of the new state, which determines the potential of future payoffs. In particular, when choosing his actions, each player has to balance these forces, a decision that may often be difficult. Although this dichotomy is also present in one-player sequential decision problems, the presence of additional players who maximize their own goals adds complexity to the analysis of the situation.

Shapley's paper in PNAS introduced the model of stochastic games with positive stopping probabilities, assuming a finite set of states, and a finite set of possible actions for each player in each state. In this model, the current state and the players' actions determine the probability that the game will terminate once the current stage is over. His paper deals with two-player zero-sum games, so that the gain of player 1 is always equal to the loss of player 2.

A history of length t in a stochastic game is the sequence of states that the game visited in the first t stages, as well as the actions that the players played in the first $t − 1$ stages. A strategy of a player is a prescription how to play the game; that is, a function that assigns to every finite history an action to play should that history occur. A behavior strategy of a player is a function that assigns to every finite history a lottery over the set of available actions.

A zero-sum game is defined to have a "value" v if (i) player 1 has a strategy (which is then said to be "optimal"), which ensures that his expected overall payoff over time does not fall below v, no matter what is the strategy followed by player 2, and (ii) if the symmetric property holds when exchanging the roles of the two players. Shapley proved the existence of a value. Because the parameters that define the game are independent of time, the situation that the players face if today the play is in a certain state is the same situation they would face tomorrow if tomorrow the play is in that state. In particular, one expects to have optimal strategies that are stationary Markov, that is, they depend only on the current state of the game. Shapley proved that indeed such optimal strategies exist, and characterized the value as the unique fixed point of a nonlinear functional operator—a two-player version of the dynamic programming principle.

The existence of stationary Markov optimal strategies implies that, to play well, a player needs to know only the current state. In particular, the value of the game does not change if players receive partial information

on each other's actions, and/or if they forget previously visited states.

In the following paragraphs, we briefly describe some of the directions in which the theory has developed following Shapley's groundbreaking paper.

2. Discounted Stochastic Games

Shapley's model is equivalent to one in which players discount their future payoffs according to a discount factor that depends on the current state and on the players' actions. The game is called "discounted" if all stopping probabilities equal the same constant, and one minus this constant is called the "discount factor." Models of discounted stochastic games are prevalent in economics, where the discount factor has a clear economic interpretation.

A large literature followed, relaxing both the zero-sum and the finiteness assumptions of Shapley's paper. In non–zero-sum games, a collection of strategies, one for each player, is a "(Nash) equilibrium" if no player can profit by deviating from his strategy, assuming all other players follow their prescribed strategies.

The stationary structure leads one to expect that equilibria in stationary Markov strategies do exist. This is indeed the case when there are finitely many states [Fink (9) and Takahashi (10)] or in the model with infinitely many states under some restrictions on the transitions [see, e.g., Parthasarathy and Sinha (11) and Nowak (12)]. However, in general, stationary Markov equilibria might not exist [Simon (13), Levy (14), and Levy and McLennan (15)].

The fact that a stationary equilibrium exists has several implications. Stationary Markov strategies are conceptually straightforward: past play affects the players' future behavior only through the current state, that is, bygones are bygones. Moreover, although a player may have deviated from equilibrium behavior in the past, thereby affecting past play, the deviation does not affect the way players play in the future. In fact, irrespective of past play, the strategies induced in the continuation game form an equilibrium of this continuation game; that is, equilibrium behavior does not involve noncredible threats, a property that is stronger than equilibrium property, and viewed as highly desirable [see Selten (16)].

3. Nonterminating Stochastic Games and Robust Solutions

Shapley assumed exogenously given positive terminating probabilities. Shortly afterward, in the series on game theory published by the Princeton University Press in the 1950s, a number of papers started investigating what happens when this assumption is dropped. Milnor and Shapley (17) studied "ruin games," in which two players, each endowed with some

initial wealth, repeatedly play a zero-sum game that affects their wealth, until one player is ruined and declared loser. Everett (18) studied the so-called "recursive games," in which termination probabilities are endogenous, and a payoff is only received upon termination of the game, as a function of the terminating state. Gillette (19) introduced a deceptively simple-looking game, inspired by playground practice, whose analysis resisted all attempts until Blackwell and Ferguson (20) provided its solution.

In a different but related framework, Aumann, Maschler, and Stearns (ref. 21, reprinted in 1995; see pp. 130–139) conceptualized a notion of a uniform solution to infinite-horizon games, in which payoffs are received in each round. This approach views the infinite-horizon game as an idealized model for long-run interactions, in which the duration of the game is sufficiently long and/or the discount factor is sufficiently close to 1, and both are not perfectly known. It aims at providing solutions whose optimality properties are not too sensitive to the exact specification of duration/discount factors. Such a (zero-sum) game is said to have a "uniform value v" if (i) the values v_T of the T-round game converge to v as the duration T of the game increases to infinity and (ii) given any $\varepsilon > 0$, each player has a strategy which is ε-optimal in all T-round games, as soon as T is large enough. The latter property implies that this strategy is also approximately optimal in discounted games, provided players are patient enough, that is, the discount factor is sufficiently close to 1. This notion of uniform ε-optimality, although giving up on exact optimality, captures robust optimality requirements.

Bewley and Kohlberg (22) proved that, in any stochastic game, the values v_T converge as the duration T increases. Mertens and Neyman (23, 24) proved that stochastic games (with a finite set of states and finite action sets) have a uniform value. The ε-optimal strategy that Mertens and Neyman constructed is complex yet allows for a simple intuition. At each round, the strategy updates a fictitious discount factor and plays an optimal strategy for that fictitious parameter. This parameter summarizes past play, and its updating is based on past payoffs. If payoffs received so far are high, the player puts higher weight on the future and increases his patience; that is, lets the fictitious discount factor get closer to 1. When the total payoff that he received so far is low, he should focus more about short-term payoffs, and therefore decrease this fictitious discount factor. The complexity of such a policy lies in the fine-tuning of the updating rule, to avoid long-term fluctuations. This construction hinges

on algebraic properties of the value of the discounted game, as a function of the discount factor, proven by Bewley and Kohlberg (22). One can view the monetary policy of the central bank, that of changing the interest rate from time to time, as an implementation of this type of strategy.

The theory of the non–zero-sum case has flourished in the last two decades but has not been settled yet. A collection of strategies, one for each player, is a uniform ε-equilibrium if by deviating no player can profit more than ε, provided the game is sufficiently long or the discount factor sufficiently close to 1, and if the payoff vector is almost independent of the duration of the game. Existence of uniform ε-equilibria has been established in few cases: Vieille (25, 26) for two-player games, Solan (27) for three-player games in which the state of the game changes at most once, and Solan and Vieille (28), Flesch et al. (29), and Simon (30) in other classes of stochastic games. The question in its most generality remains open.

A weaker concept than Nash equilibrium is that of "correlated equilibrium," introduced by Aumann (31, 32). A collection of strategies, one for each player, is a correlated equilibrium if it is a Nash equilibrium in an extended game that contains a correlation device, which sends signals to the players along the play. The correlation device may send public or private signals, at the outset of the game or at the beginning of every stage, and the signals can be independent or correlated. Each of these variations gives rise to another concept of correlated equilibrium. Nowak and Raghavan (33) proved the existence of a correlated equilibrium in discounted stochastic games, in which the strategies of the players are stationary Markov and the signals of the correlation device are public. Building upon Mertens and Neyman (23), Solan and Vieille (34) proved the existence of a correlated uniform ε-equilibrium in which the correlation device sends a private message to each player at every round, which is correlated with the messages sent in the previous round.

Another strand of literature studies stochastic games in continuous time, and their relation to discrete-time approximations of the game. It turns out that the correlated uniform ε-equilibrium of Solan and Vieille (34) can be transformed into a Nash equilibrium in the continuous-time game [see Neyman (35)].

In the zero-sum case, other evaluations have attracted interest. Rather than viewing the infinite-horizon game as a stylized model of long-run interactions, one may instead assign to each infinite play of the game a payoff, which summarizes payoffs received along the play. Player 1 (respectively, player 2) maximizes (respectively, minimizes) the expectation

of this payoff. When the payoff of a play is defined to be the upper limit of the averages of the stage payoffs received along the play, a value was shown to exist when the state space is general by Maitra and Sudderth (36). Martin (37) proved a far-reaching result, establishing the existence of the value as soon as the payoff function satisfies minimal (measurability) requirements.

4. A General Model of Repeated Games

In the 1960s, commissioned by the United States, Aumann, Maschler, and Stearns initiated and developed the theory of (two-player, zero-sum) repeated games with incomplete information. In such games, the state is chosen at the outset of the game, kept fixed throughout the play, and the two players have imperfect and possibly asymmetric knowledge of it. The main goal of the analysis is to understand to what extent privately held information is valuable in the long run, and how to make optimal use of it. Although different from stochastic games in some important dimensions, the two theories were shown to share common features. After some developments, not surveyed here [see Zamir (38)], a general model of repeated/stochastic games with signals was introduced by Mertens (39). Such games have a finite number of possible states, and the two players have a finite set of actions available in each state. In addition, there are two finite sets of "private signals" for the two players. In each round, the current state and the actions being played determine the probability distribution according to which the next state of the game and the signals to be privately provided to the two players are chosen. These signals are all that the players get to know along the play. This unified model is flexible enough to include stochastic games; repeated games with incomplete information, imperfect monitoring, or information lags [Scarf and Shapley (40)]; and many other models, such as game-theoretic versions of partially observed Markov decision processes. It is worth noting that the general model of repeated games is a stochastic game with imperfect monitoring of the state and actions.

The mathematical study of this general model is currently very active [see Laraki and Sorin (41) for a recent review of the topic].

5. Applications

Stochastic games provide a model for a large variety of dynamic interactions and are therefore useful in modeling real-life situations that arise in, e.g., economics, political science, and operations research. So that the analysis of the game provides decisive predictions and recommendations, the data that define it must have special features. Because applications are usually motivated by the search for clear-cut conclusions, only highly structured models of stochastic games have been studied.

The significance of stochastic games is threefold. First, by modeling a dynamic situation as a stochastic game, researchers must understand the structure of the problem they face. Second, to simplify the model, they have to realize which aspects of the model do not affect the outcome and can be dispensed with. Third, the qualitative predictions of the model sometimes provide useful conclusions. We provide here two applications of stochastic games.

One area that was extensively studied as a stochastic game is the overexploitation of a common resource, which goes back to Lloyd (42). Levhari and Mirman (43) studied a fishery war between two countries. The state variable is the quantity of fish in a given area, which grows exponentially in the absence of unnatural intervention. Each one of two countries has to determine the quantity of fish it allows its fishermen to catch, so as to maximize its long-run utility. The authors concluded that, in equilibrium, the fish population will be smaller than the population that would have resulted if the two countries cooperated and maximized their joint utility. The phenomenon of overexploitation of a common resource is known in economics as the "tragedy of the commons." A complete characterization of the set of equilibria in this model has been provided by Chiarella et al. (44) and Dutta and Sundaram (45), who found out that there may be equilibria in which the common resource is underexploited, so that the tragedy of the commons need not occur.

A second application of stochastic games is that of market games with money. In a sequence of papers, Karatzas et al. (46, 47) and Geanakoplos et al. (48) studied the origin of inflation in a market game with continuum of agents and a central bank. At every stage, each player receives a random endowment of a perishable commodity, decides how much to lend to or to borrow from the central bank, and consumes the amount that he has after this transaction. The conclusion of the analysis is that the mere presence of uncertainty in the endowments leads to inflation.

6. Shapley's Contribution in Perspective

Shapley's work on stochastic games, which was published 60 y ago, paved the way to a variety of dynamic models of games that proved fruitful in many areas and have been the subject of an extremely rich and diverse research in the last 50 y [see, e.g., Neyman and Sorin (49) and Mertens, Sorin, and Zamir (50)]. A hoard of new tools has been developed to handle questions that were impenetrable in Shapley's time, new connections between game theory and various topics in mathematics have been established, and new areas of application have developed, mostly in economics, computer science, and operations research [see, e.g., Altman and Hordijk (51)].

Although our understanding of dynamic situations has improved, the questions that we can answer are still limited, and the models that are analyzed are still very stylistic. New tools must be developed so that we can treat models that are closer to real-life situations and provide better predictions. This is the challenge that faces us in the coming 60 y.

1 von Neumann J, Morgenstern O (1944) *Theory of Games and Economic Behavior* (Princeton Univ Press, Princeton).
2 Nash JF (1950) Equilibrium points in *n*-person games. *Proc Natl Acad Sci USA* 36(1):48–49.
3 Nash J (1950) The bargaining problem. *Econometrica* 18(2):155–162.
4 Shapley LS (1953) Stochastic games. *Proc Natl Acad Sci USA* 39(10):1095–1100.
5 Shapley LS (1953) A value for *n*-person games. *Contributions to the Theory of Games*, II, Annals of Mathematical Studies, eds Kuhn WH, Tucker AW (Princeton Univ Press, Princeton), Vol 28, pp 307–317.
6 Kuhn HW (1957) Extensive games and the problem and information. *Contributions to the Theory of Games*, II, Annals of Mathematical Studies, eds Kuhn HW, Tucker AW (Princeton Univ Press, Princeton), Vol 28, pp 193–216.
7 Luce RD, Raiffa H (1957) *Games and Decisions: Introduction and Critical Survey* (Wiley, New York).
8 von Neumann J (1928) Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen* 100(1):295–320; trans Tucker AW, Luce RD, eds (1959) [*Contribution to the Theory of Games*, IV, Annals of Mathematical Studies] (Princeton Univ Press, Princeton), Vol 40. German.
9 Fink AM (1964) Equilibrium in a stochastic *n*-person game. *J Sci Hiroshima Univ* 28(1):89–93.

10 Takahashi M (1964) Equilibrium points of stochastic non-cooperative *n*-person games. *J Sci Hiroshima Univ Ser A-I Math* 28(1):95–99.
11 Parthasarathy T, Sinha S (1989) Existence of stationary equilibrium strategies in non-zero sum discounted stochastic games with uncountable state space and state-independent transitions. *Int J Game Theory* 18(2):189–194.
12 Nowak A (2003) On a new class of non-zero-sum discounted stochastic games having stationary nash equilibrium points. *Int J Game Theory* 32(1):121–132.
13 Simon RS (2003) Games of incomplete information, ergodic theory, and the measurability of equilibria. *Israel J Math* 138(1):73–92.
14 Levy J (2013) A discounted stochastic game with no stationary Nash equilibrium: Two examples. *Econometrica* 81(5):1973–2007.
15 Levy J, McLennan A (2015) Corrigendum to "A discounted stochastic game with no stationary Nash equilibrium: Two examples." *Econometrica* 83(3):1237–1252.
16 Selten R (1975) Reexamination of the perfectness concept for equilibrium points in extensive games. *Int J Game Theory* 4(1):25–55.
17 Milnor J, Shapley LS (1991) On games of survival. *Stochastic Games and Related Topics*, eds Raghavan TES, Ferguson TS, Parthasarathy T, Vrieze OJ (Springer, Dordrecht, The Netherlands), pp 207–233.

18 Everett H (1957) Recursive games. *Contributions to the Theory of Games*, III, Annals of Mathematical Studies, eds Dresher M, Tucker AW, Wolfe P (Princeton Univ Press, Princeton), Vol 39, pp 47–78.

19 Gillette D (1957) Stochastic games with zero stop probabilities. *Contributions to the Theory of Games*, III, Annals of Mathematical Studies, eds Dresher M, Tucker AW, Wolfe P (Princeton Univ Press, Princeton), Vol 39, pp 179–187.

20 Blackwell D, Ferguson TS (1968) The big match. *Ann Math Stat* 39(1):159–163.

21 Aumann RJ, Maschler M (1995) *Repeated Games with Incomplete Information* (MIT Press, Cambridge, MA).

22 Bewley T, Kohlberg E (1976) The asymptotic theory of stochastic games. *Math Oper Res* 1(3):321–336.

23 Mertens J-F, Neyman A (1981) Stochastic games. *Int J Game Theory* 10(2):53–66.

24 Mertens J-F, Neyman A (1982) Stochastic games have a value. *Proc Natl Acad Sci USA* 79(6):2145–2146.

25 Vieille N (2000) Equilibrium in 2-person stochastic games I: A reduction. *Israel J Math* 119(1):55–91.

26 Vieille N (2000) Equilibrium in 2-person stochastic games II: The case of recursive games. *Israel J Math* 119(1):93–126.

27 Solan E (1999) Three-person absorbing games. *Math Oper Res* 24(3):669–698.

28 Solan E, Vieille N (2001) Quitting games. *Math Oper Res* 26(2):265–285.

29 Flesch J, Kuipers J, Schoenmakers G, Vrieze K (2010) Subgame-perfection in positive recursive games with perfect information. *Math Oper Res* 35(1):193–207.

30 Simon RS (2015) The challenge of non-zero-sum stochastic games. *Int J Game Theory*, 10.1007/s00182-015-0497-3.

31 Aumann RJ (1974) Subjectivity and correlation in randomized strategies. *J Math Econ* 1(1):67–96.

32 Aumann RJ (1987) Correlated equilibrium as an expression of Bayesian rationality. *Econometrica* 55(1):1–18.

33 Nowak A, Raghavan TES (1992) Existence of stationary correlated equilibria with symmetric information for discounted stochastic games. *Math Oper Res* 17(3):519–526.

34 Solan E, Vieille N (2002) Correlated equilibrium in stochastic games. *Games Econ Behav* 38(2):362–399.

35 Neyman A (2012) Continuous-time stochastic games. *Discussion Paper Series* (The Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Jerusalem), no. dp616.

36 Maitra AP, Sudderth WD (1993) Borel stochastic games with limsup payoffs. *Ann Probab* 21(2):861–885.

37 Martin DA (1998) The determinacy of Blackwell games. *J Symb Log* 63(4):1565–1581.

38 Zamir S (1992) Repeated games of incomplete information: Zero-sum. *Handbook of Game Theory*, eds Aumann RJ, Hart S (North Holland, Amsterdam), Vol 1, pp 109–154.

39 Mertens J-F (1987) Repeated games. *Proceedings of the International Congress of Mathematicians, Berkeley, 1986* (American Mathematical Society, Providence, RI), pp 1528–1577.

40 Scarf HE, Shapley LS (1957) Games with partial information. *Contributions to the Theory of Games*, III, Annals of Mathematical Studies, eds Dresher M, Tucker AW, Wolfe P (Princeton Univ Press, Princeton), Vol 39, pp 213–229.

41 Laraki R, Sorin S (2015) Advances in zero-sum dynamic games. *Handbook of Game Theory*, eds Young HP, Zamir S (North Holland, Amsterdam), Vol 4, pp 27–95.

42 Lloyd WF (1833) *Two Lectures on the Checks to Population* (University of Oxford, Oxford).

43 Levhari D, Mirman L (1980) The Great Fish War: An example using a dynamic Cournot-Nash solution. *Bell J Econ* 11(1):322–334.

44 Chiarella C, Kemp MC, Long NV, Okuguchi K (1984) On the economics of international fisheries. *Int Econ Rev* 25(1):85–92.

45 Dutta PK, Sundaram RK (1993) The tragedy of the commons? *Econ Theory* 3(3):413–426.

46 Karatzas I, Shubik M, Sudderth WD (1997) A strategic market game with secured lending. *J Math Econ* 28(2):207–247.

47 Karatzas I, Shubik M, Sudderth WD, Geanakoplos J (2006) The inflationary bias of real uncertainty and the harmonic Fisher equation. *Econ Theory* 28(3):481–512.

48 Geanakoplos J, Karatzas I, Shubik M, Sudderth WD (2014) Inflationary equilibrium in a stochastic economy with independent agents. *J Math Econ* 52:1–11.

49 Neyman A, Sorin S (2003) *Stochastic Games and Applications* (Kluwer Academic Publishers, Dordrecht, The Netherlands), Vol 570.

50 Mertens J-F, Sorin S, Zamir S (2015) *Repeated Games* (Cambridge Univ Press, Cambridge, UK).

51 Altman E, Hordijk A (1995) Zero-sum Markov games and worst-case optimal control of queueing systems. *QUESTA* 21(3):415–447.