

MIT OpenCourseWare
<http://ocw.mit.edu>

24.222 Decisions, Games, and Rational Choice
Spring 2008

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.

Notes on Games and Models for Games

A game in normal or strategic form may be defined formally by specifying the following elements: (1) a finite set N , the *players*; (2) for each player i , a non-empty set, C_i , representing the *strategies* that are available to that player; (3) for each player i , a function U_i , taking *strategy profiles* (sequences of one strategy for each player) to real numbers, representing the *utility value* of the outcome of the game that results from the sequence of strategy choices given by the profile. A simple two-person game can be represented in a two-dimensional matrix, as in figure one, with a row for each strategy in the first player's strategy set, and a column for each strategy in the second player's strategy set. Each cell of the matrix represents a strategy profile (or pair, in the two-person case), and will contain two real numbers, representing the utilities of the two players for the outcome that results from that strategy pair.

A game is only a *partial* specification of the decision problems faced by the players, since it specifies only the utilities that motivate the players, and not their beliefs about the choices of the other players. But we may make qualitative assumptions that constrain the players' beliefs about each other, for example, by assuming that players are rational, and that it is common knowledge or common belief among the players that they all are rational. To make such assumptions precise, and to determine exactly what follows from the assumptions about what the players will do, we need to complete the specification of a particular playing of the game by players with particular beliefs and degrees of belief, both about what the other players will do, and what the other players believe about each other. The further specification is given by a *model* of a game.

Formally a model for a game is a structure, $\langle W, \mathbf{a}, P, S \rangle$ where

W is a set of possible worlds (the state space) and \mathbf{a} is a designated member of W representing the actual world of the model.

P is a function determining for each $i \in N$ and $x \in W$ a conditional probability function. Where ϕ and ψ are propositions, or (to use the probability theorist's terminology) *events*, $P_{i,x}(\phi/\psi)$ represents the beliefs player i has at the start of the game in possible world x . " $P_{i,x}(\phi/\psi) = r$ " says that at the start of the game as it is played in possible world x , player i 's degree of belief in proposition ϕ conditional on proposition ψ is r . The conditional probability functions are defined for all propositions, including the case where the probability of the condition is zero. The conditional probabilities represent, in such cases, the player's dispositions to revise her beliefs upon being surprised, where a surprise is the discovery that something with prior probability 0 was true. In the case where the condition is the empty, or impossible, proposition, the conditional probability of all propositions is stipulated to be 1.

S is a function taking the possible worlds to strategy profiles. So, for player i , $S_i(x)$ will be the strategy played by that player in possible world x . $S(x)$, without the subscript, will be the full

sequence or profile of strategies, one for each player.

It is useful to define an equivalence relation, for each player, on the possible worlds of the model, a relation of *subjective indistinguishability*. Two possible worlds are subjectively indistinguishable, to player i , just in case player i 's beliefs and strategy choice are exactly the same in two worlds. The intuitive idea is that from the players perspective (at the start of the game), the two worlds seem exactly the same. The definition is simple: x is subjective indistinguishable from y to player i if and only if $P_{i,x}(\phi/\psi) = P_{i,y}(\phi/\psi)$ for all ϕ and ψ . A player's *type* in a given possible world x , $t(x)$, will be the equivalence class of world x in which the player has exactly that total belief state. We assume that there are enough possible worlds in our model so that any type for one player is compatible with any configuration of types for the others. It is also assumed that players know their own type (they know what they themselves believe, and what strategy choice they themselves make.)

A *proposition*, in a model, is a subset of W . Intuitively, it is the set of possible states of the world in which the proposition is true. So, for example, one can define the possible worlds in which all the players are rational—the worlds in which the choices they make in that world maximize expected utility, given the beliefs and degrees of belief that they have in that world. For any proposition, one can define the proposition that a player *believes* that proposition—the worlds in which the player's probability function assigns one to the proposition. One can define a notion of *common belief* between two or more players: a group of players have common belief that ϕ if all believe that ϕ , all believe that all believe it, all believe that all believe that all believe it, . . . (infinitely iterated). So specifically, one can define the proposition that there is common belief among the players that all of them are rational. Then one can define the class of strategies that are played in the actual world of some model in which the players all have this belief, and act rationally on it. These will be just the strategies that survive the algorithm, iterated elimination of strictly dominated strategies.

The algorithm of iterated elimination of strictly dominated strategies is easy to state and apply. A strategy is *strictly dominated* if there is some alternative strategy (pure or mixed) that has greater utility relative to every strategy profile for the other players. For any game, we can define a reduced game that results from eliminating every strategy that is strictly dominated. Then one can apply the same elimination procedure to the reduced game, and continue this until the resulting game is one in which no strategy for any player is strictly dominated.

The only twist in applying this algorithm is that one must be sure to take account of strategies that are dominated by mixed strategies (a mixed strategy is a probability distribution over some subset of a player's strategies). Consider, for example, the game in figure two. Strategy A is not strictly dominated by any pure strategy, but is strictly dominated by a 50/50 mix of strategies B and C. As a result, no matter what Player I believes about what player II will do, it cannot be rational for her to choose A. If she thinks X is more likely, she should choose C, and if she thinks Y is more likely, she should choose B. If X and Y have equal probability for her, then she will be indifferent between B and C, but both will be better than A.

A strategy is *weakly dominated* if there is some alternative strategy (pure or mixed) that has

equal or greater utility relative to every strategy profile for the other players, and *greater* utility for at least one such profile. A more restrictive algorithm (one that eliminates more strategies) is the iterated elimination of *weakly* dominated strategies. If the algorithm is applied to figure 3, the money burning game, the only strategies that survive are \sim BU for player I and LL for player II. The elimination of weakly dominated strategies is more controversial, and can be justified only with stringent, and somewhat contrived epistemic conditions.

Figure one

		Player II		
		c_{II}	d_{II}	
Player I	c_I	0	9	5
	d_I	1	1	0

$N = \{I, II\}$

$C_I = \{c_I, d_I\}$

$C_{II} = \{c_{II}, d_{II}\}$

Figure two

		Player II	
		X	Y
Player I	A	2	5
	B	1	5
	C	5	1

Figure three

		LL	LR	RL	RR
BU	2	1	2	-2	-2
BD	-2	0	0	-1	-1
~BU	4	1	0	4	0
~BD	0	0	4	0	4