

Partial Reversible Game Theory

by Sven Nilsen, 2018

In traditional game theory, the whole point is that rewards are irreversible – the players get to keep the reward after receiving it. In a total reversible game with repeated plays, the optimal strategy can be found simply by winding back time and try something at random until the received reward is maximum. However, there is an interesting kind of game between irreversible and reversible: A partial reversible game. In this paper I represent the necessary structure to describe such games and explain why these games are strictly more complex than in traditional game theory.

In a simple irreversible game with two players A and B, the rewards can be described as a 2x2 matrix:

A / B	A1	A2
B1	10 / 5	-4 / 8
B2	8 / 3	-3 / 2

Each cell in the matrix contains the rewards `A / B` which the players receive after making their choice. There must be at least two choices per player, otherwise the game would not be very interesting. Also, the players get to keep the rewards, but they also must pay the price if the reward is negative.

Such games has been studied extensively for many years.

In a partial reversible game, the reward is split into two components per player: One part which the player keeps no matter whether the game is reversed or not. The second part is kept only if the game is NOT reversed. A such reward matrix might look as following (same total payoffs as before):

A / B (irreversible, reversible)	A1	A2
B1	5, 5 / 3, 2	-3, -1 / 2, 6
B2	7, 1 / 3, 0	-2, -2 / 1, 1

What makes such games strictly more complex is that a rational choice depends on the condition for which the game is reversed. This condition can be *arbitrary*.

For example, two players can agree upon reversing the game if both players agree it should be reversed later (AND). Or, they could agree upon reversing the game if either wants to reverse it later (OR).

The game could even be reversed by some random bit which the players receive partial information about before playing. This means the reversible mechanism is outside the players control, but they still have to collect evidence in advance to make a rational decision.

In traditional game theory, a rational decision is based only on the information that is contained within the reward matrix. The model of the world required to be rational is very small.

Under a partial reversible game, the model of world required to be rational might be infinitely complex.