```
function [V] =
ex_4_5_policy_evaluation(V,pol_pi,emp_pol_pi,gamma,Ra,Pa,Rb,Pb,max_num_cars_can_transfer,
max_cars_can_store)
% JCR_POLICY_EVALUATION - Performs policy evaluation evaluations returning the
%  state-value function for the jacks car rental example.
%
% The Basic Algorithm is: Iterate the Bellman equation:
%
% V(s) <- \sum_a \pi(s,a) \sum_{s'} P_{s,s'}^a (R_{s,s'}^a + \gamma V(s'))
%
% where the policy is given as input.  These iterations are done IN PLACE.
%
% See ePage 262 in the Sutton book.
%
% Input:
%   V = An array to hold the values of the state-value function
%
% Written by:
% --
% John L. Weatherwax                2007-12-03
%
% email: wax@alum.mit.edu
%
% Please send comments and especially bug reports to the
% above email address.
%
%-----

if( nargin < 3 ) gamma = 0.9; end

% the maximum number of cars at each site (assume equal):
max_n_cars = size(V,1)-1;

% the total number of states (including the states (0,Y) and (X,0)):
nStates = (max_n_cars+1)^2;

% some parameters for convergence:
%
MAX_N_ITERS = 100; iterCnt = 0;
CONV_TOL    = 1e-6;  delta = +inf;  tm = NaN;

%---
% MAIN policy evaluation loop:
%---
fprintf('beginning policy evaluation ... \n');
while( (delta > CONV_TOL) && (iterCnt <= MAX_N_ITERS) )
  delta = 0;
  % For each state in \cal{S}:
  for si=1:nStates,
    % get the number of cars (ones based) at each site (at the END of the day):
    [na1,nb1] = ind2sub( [ max_n_cars+1, max_n_cars+1 ], si );
    na = na1-1; nb = nb1-1; % (zeros based)
    %fprintf( 'prev state took = %10.5f (min); considering state =
%5d=(na=%5d,nb=%5d)...\n', tm/60, si,na,nb );

    % get the old action value for this state:
    v = V(na1,nb1);

    % tranfer this many cars from A to B according to the poliy \pi:
```

```
    ntrans = pol_pi(na1,nb1); useEmp = emp_pol_pi(na1,nb1);

    %---
    % based on the state and action compute the expectation over
    %     all possible states we may transition to i.e. s'
    % We need to consider 1) the number of possible returns at site A/B
    %                     2) the number of possible rentals at site A/B
    %---
    tic;
    V(na1,nb1) =
ex_4_5_rhs_state_value_bellman(na,nb,ntrans,useEmp,V,gamma,Ra,Pa,Rb,Pb,max_num_cars_can_t
ransfer,max_cars_can_store);
    tm=toc;

    delta = max( [ delta, abs( v - V(na1,nb1) ) ] );
  end % end state loop

  iterCnt=iterCnt+1;
  % lets print the iterations if desired:
  if( 1 && mod(iterCnt,1)==0 )
    fprintf( 'iterCnt=%5d; delta=%15.8f\n', iterCnt, delta );
    %disp( V );
    %disp( fix(V*10)/10 ); % <- just display ONE decimal
    %disp( round(V*10)/10 ); % <- just display ONE decimal
    %pause
  end
end % end while loop
fprintf('ended policy evaluation ... \n');
```