CIS 620 — Advanced Topics in AI
Profs. M. Kearns and L. Saul
A couple of (hopefully) clarifying remarks

In general, if π is any policy, then $V^\pi(s)$ denotes the expected discounted return when starting at s and following policy π. One way of computing V^π is by solving a linear system of equations in the "variables" $V^\pi(s)$, but there are other ways, and this is not part of the definition of V^π.

Now suppose I give you an arbitrary function $\hat{V}$, and we let $\hat{\pi} = greedy(\hat{V})$. There is no reason to expect $V^{\hat{\pi}}$ to equal $\hat{V}$. For instance, suppose that from some initial state s_0, the arbitrary $\hat{V}$ assigns high values to all states reachable in one step from s_0 under the "left" action, and low values to all states to the "right". The greedy policy dictated by $\hat{V}$ would of course tell us to go left; but I haven't said anything about the true rewards yet, so I could make all the left states have low rewards and all the right states have high rewards. Following policy $\hat{\pi}$ is now a bad idea, and $V^{\hat{\pi}}(s_0)$ will duly reflect this. But $\hat{V}(s_0)$ may be quite large (it's arbitrary). The point is that in Problem 2, there is no "definition" of $\hat{V}$ — just a guarantee that it's pretty close to V^* everywhere. It turns out this is enough to insure closeness, but not equality, to $V^{\hat{\pi}}$.