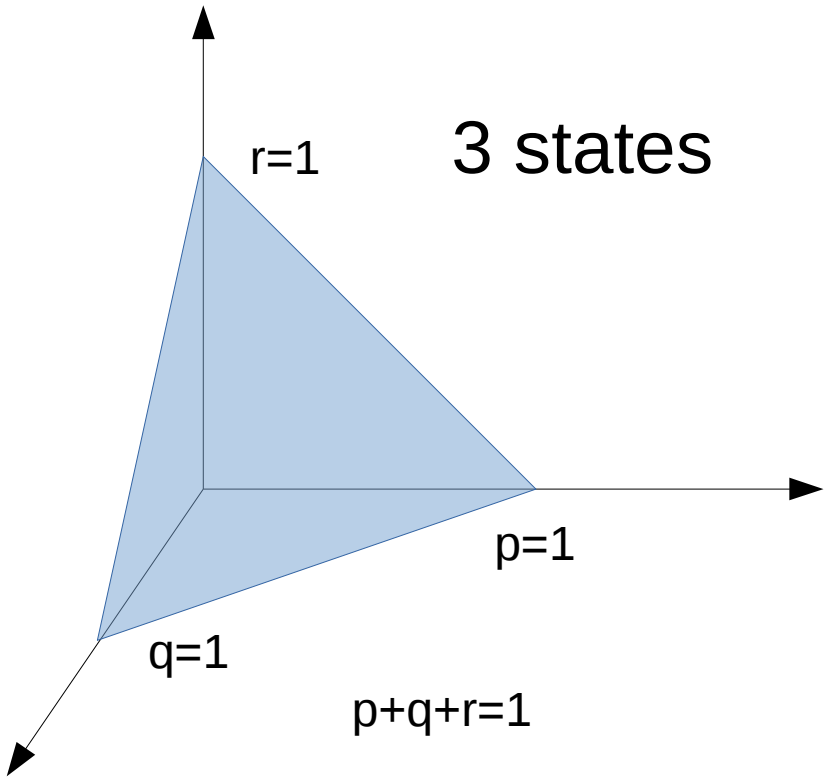
A maths-free view on POMDPs

p=0 p=1

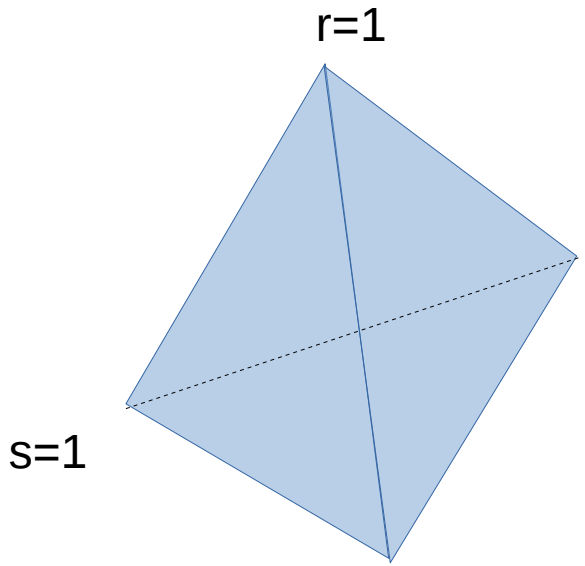
2 states

Belief states

p: belief of being in state 1 (rather than in state 0)



3 states



4 states

p+q+r+s=1

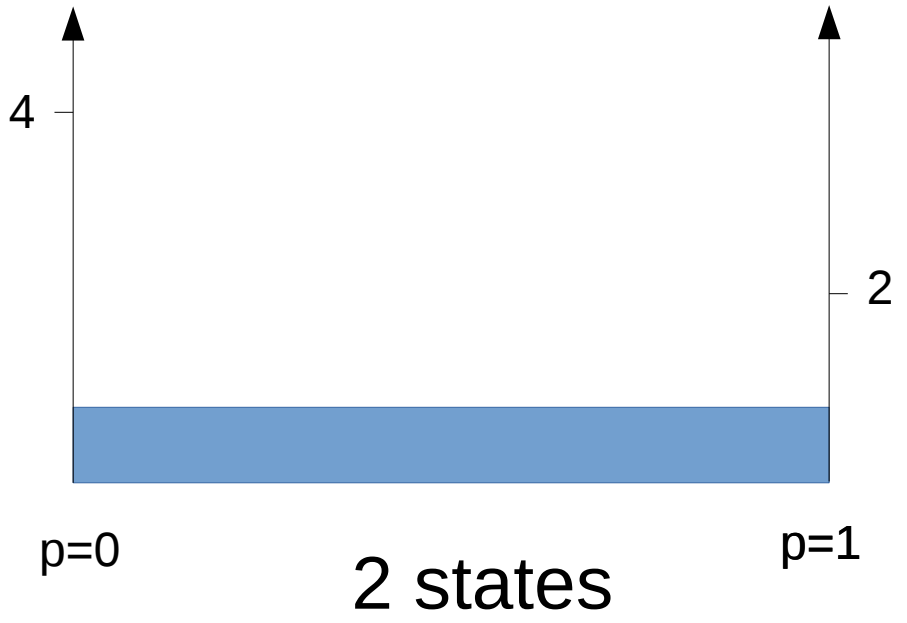
Note that the beliefs always sum up to 1, the corners indicate certainty to be in one state.

Actions over belief states

- In the example of the moving robot, with e.g. 100 grid cells, the belief space has 100 dimensions
- We can map the belief state to a distribution over the grid (as illustrated in the last lecture)
- The choice of action, value function etc. are nevertheless defined over the 100 dimensional space
- Obviously, approximations and simplifications are needed and possible
- In the following we will show a complete picture for 2 or 3 states

Actions over belief states

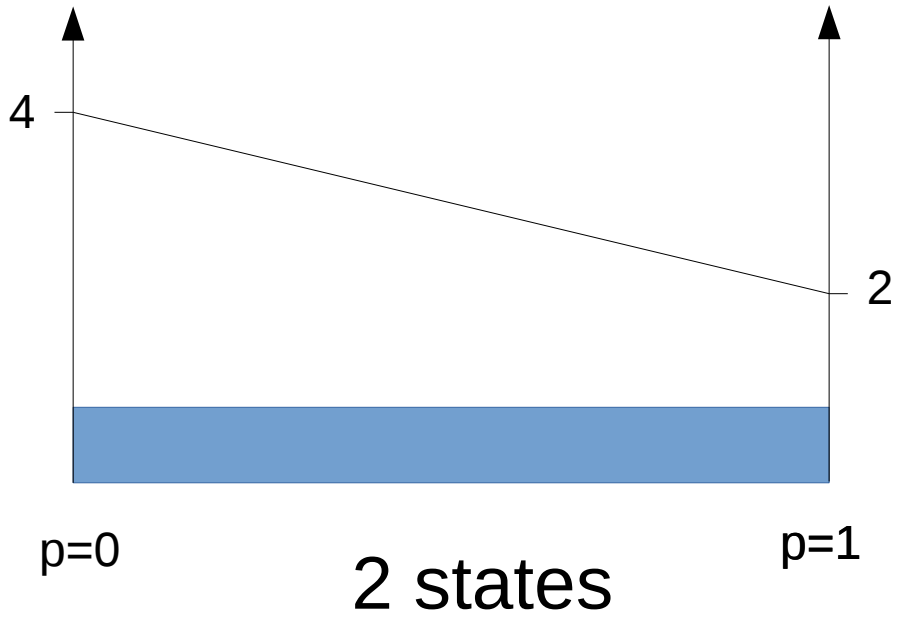
Say, action 1 leads to a reward of 4 from state 0
and to a reward of 2 from state 1



Actions over belief states

Say, action 1 leads to a reward of 4 from state 0
and to a reward of 2 from state 1

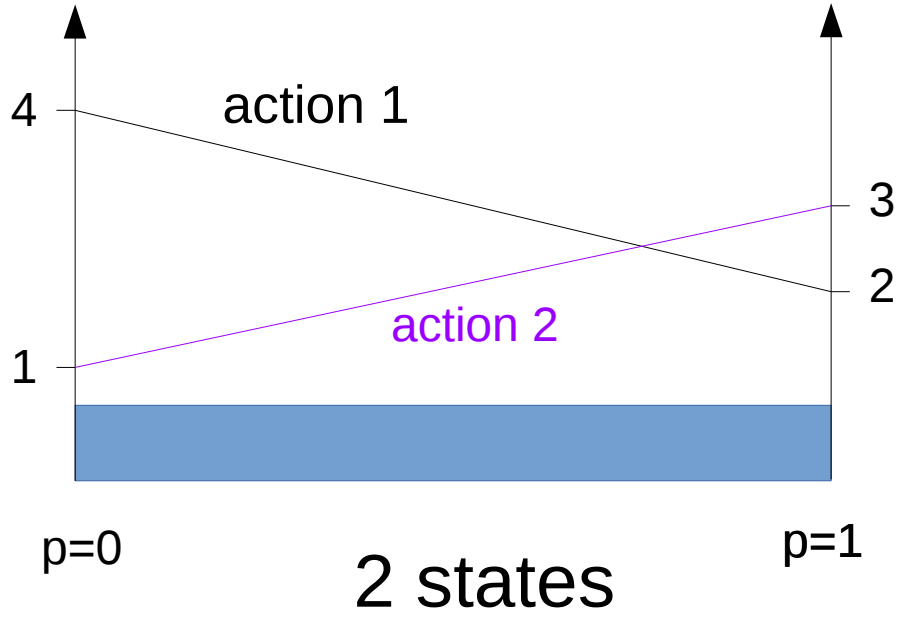
So at p=0.5 we should expect a reward of 3



Actions over belief states

Say, action 1 leads to a reward of 4 from state 0
and to a reward of 2 from state 1

Say, action 2 leads to a reward of 1 from state 0
and to a reward of 3 from state 1

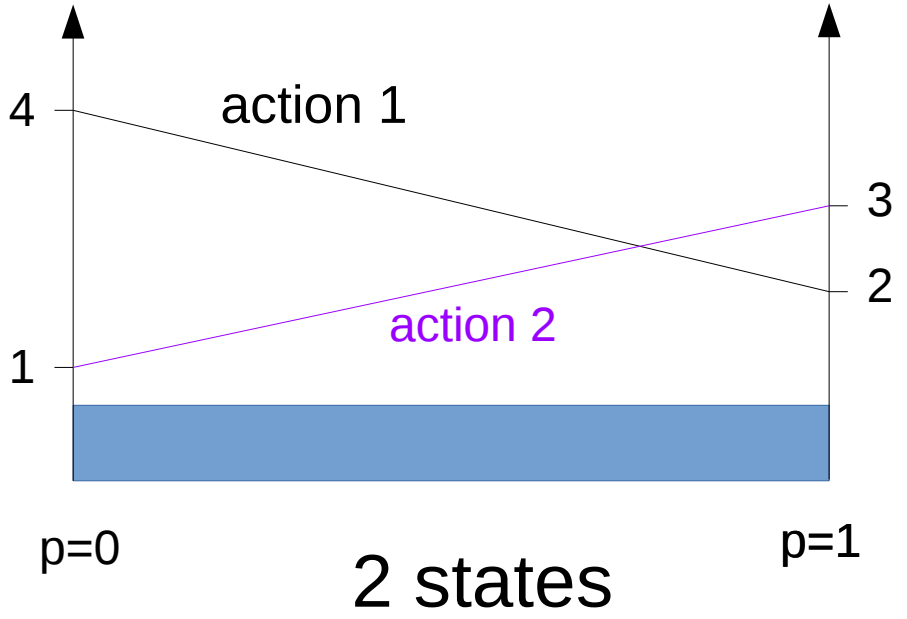


Actions over belief states

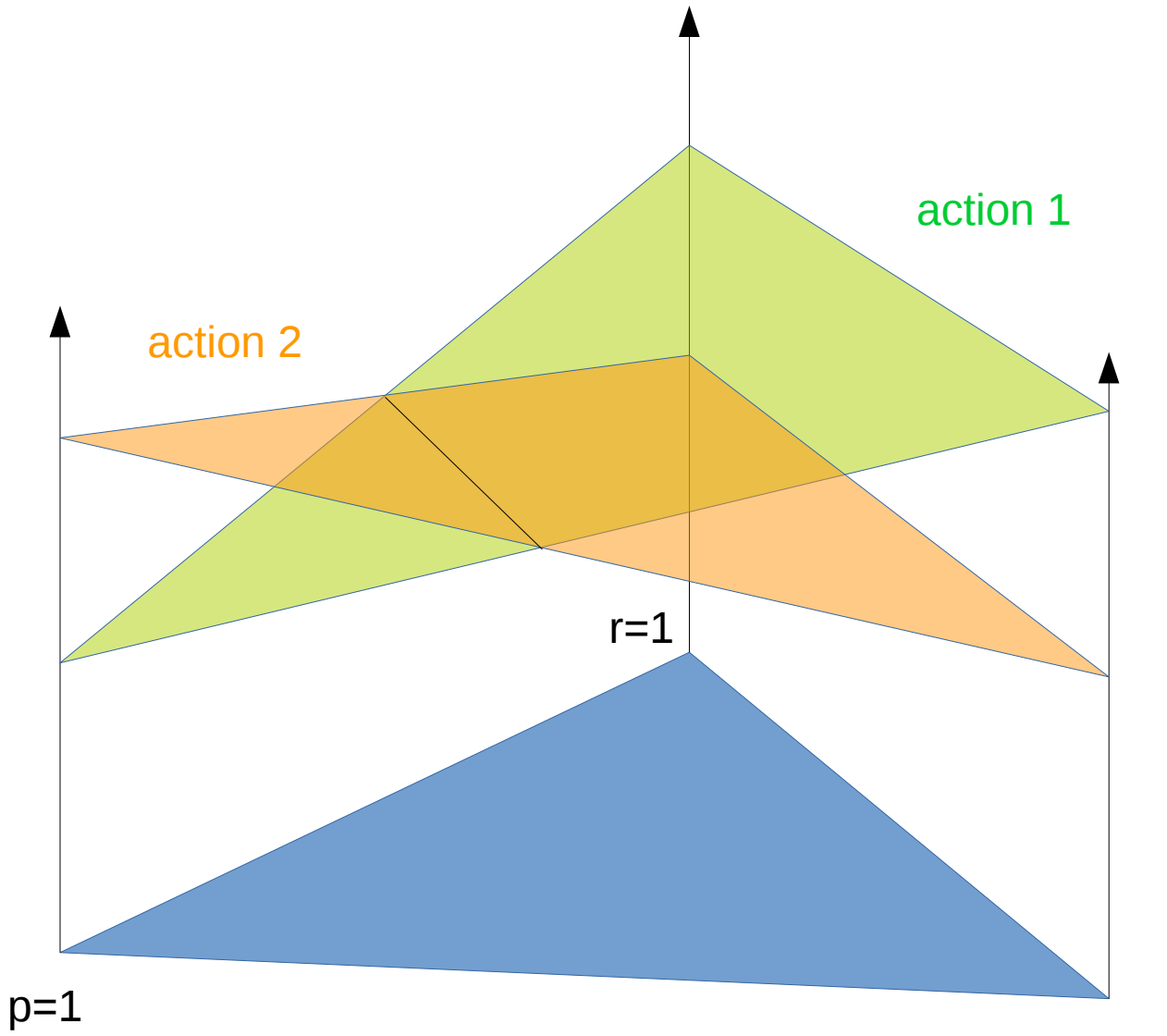
Say, action 1 leads to a reward of 4 from state 0
and to a reward of 2 from state 1

Say, action 2 leads to a reward of 1 from state 0
and to a reward of 3 from state 1

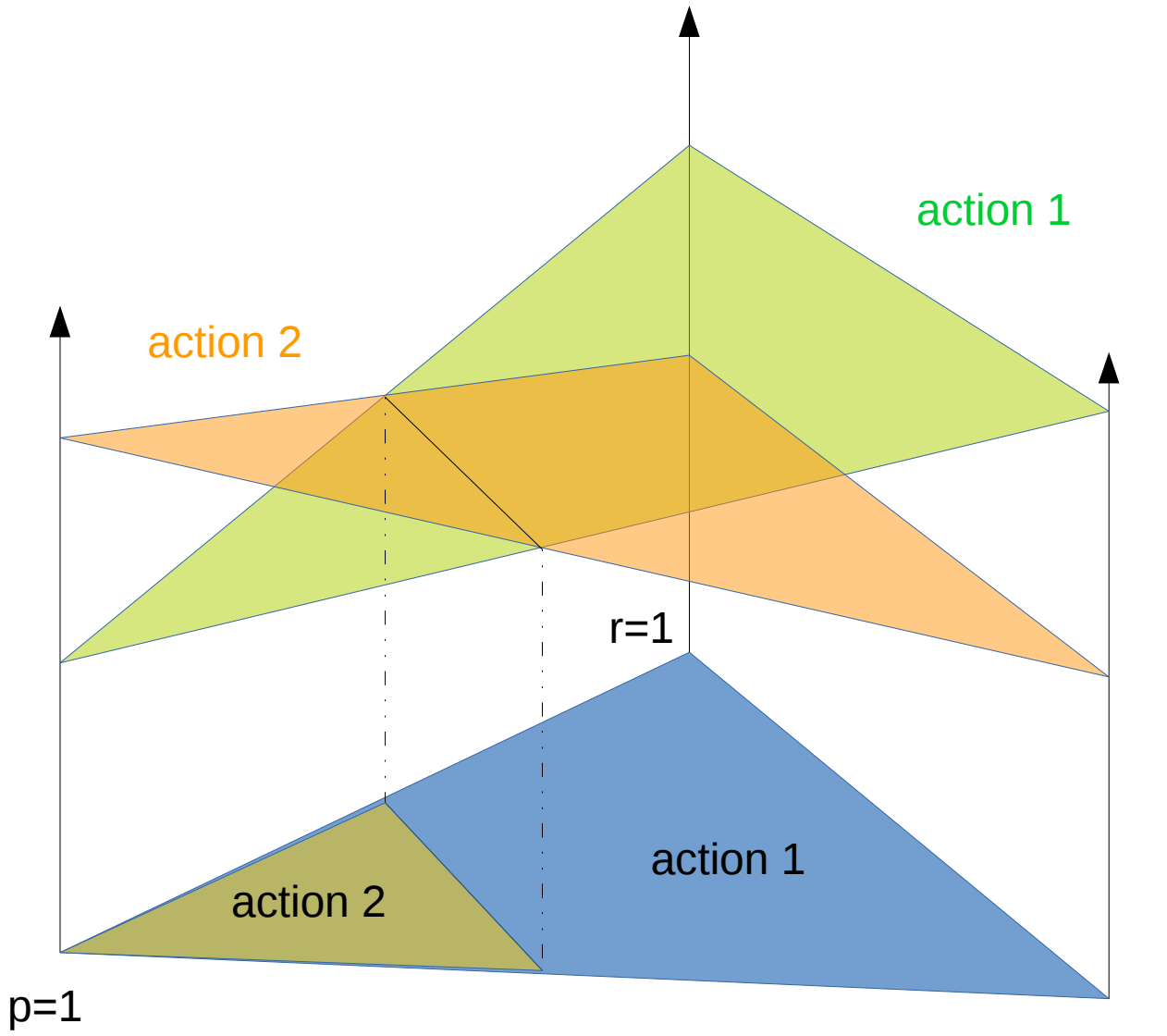
So if p<0.75 action 1 is better and if p>0.75 action 1



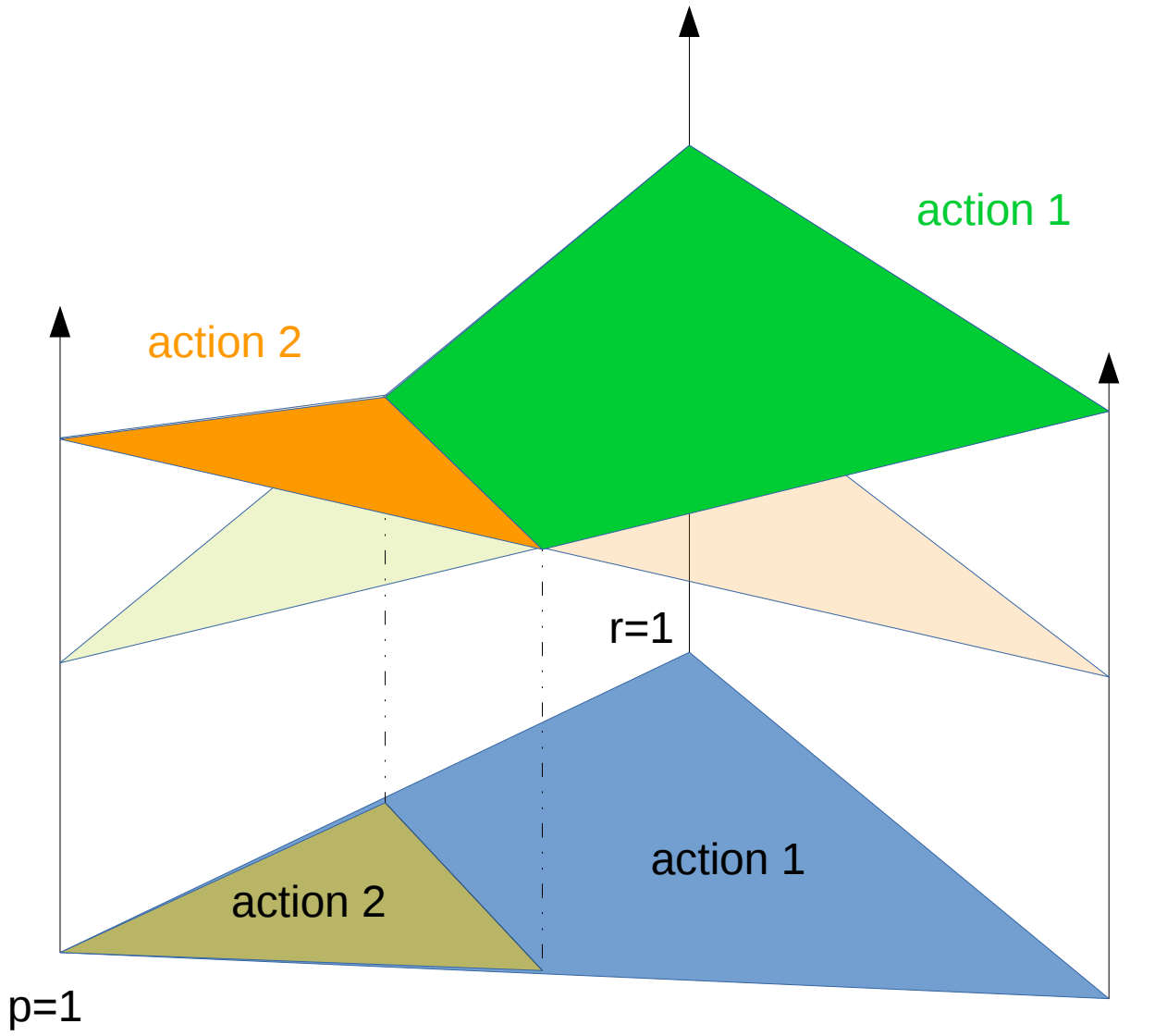
Note that the vertical axes usually give a value function (so here gamma=0)



Belief space for 3 states

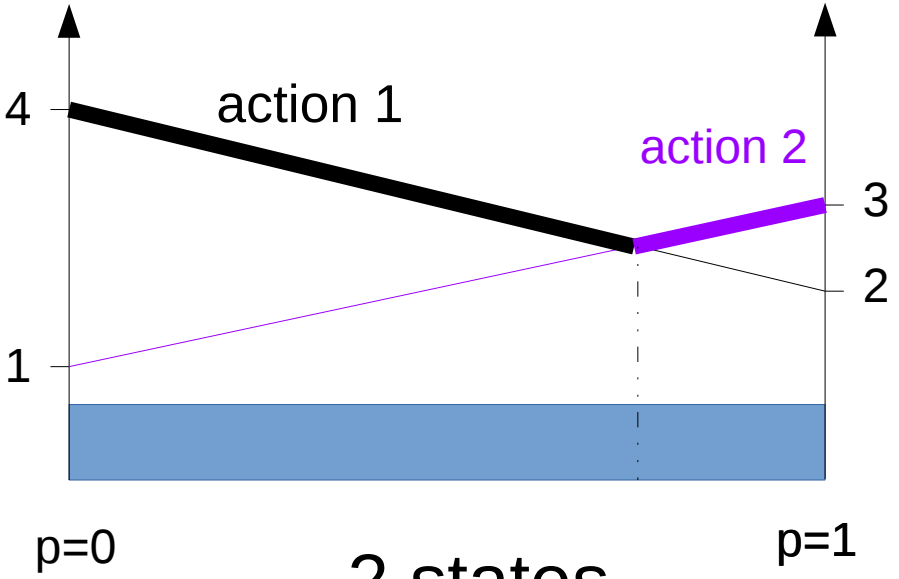


Belief space for 3 states

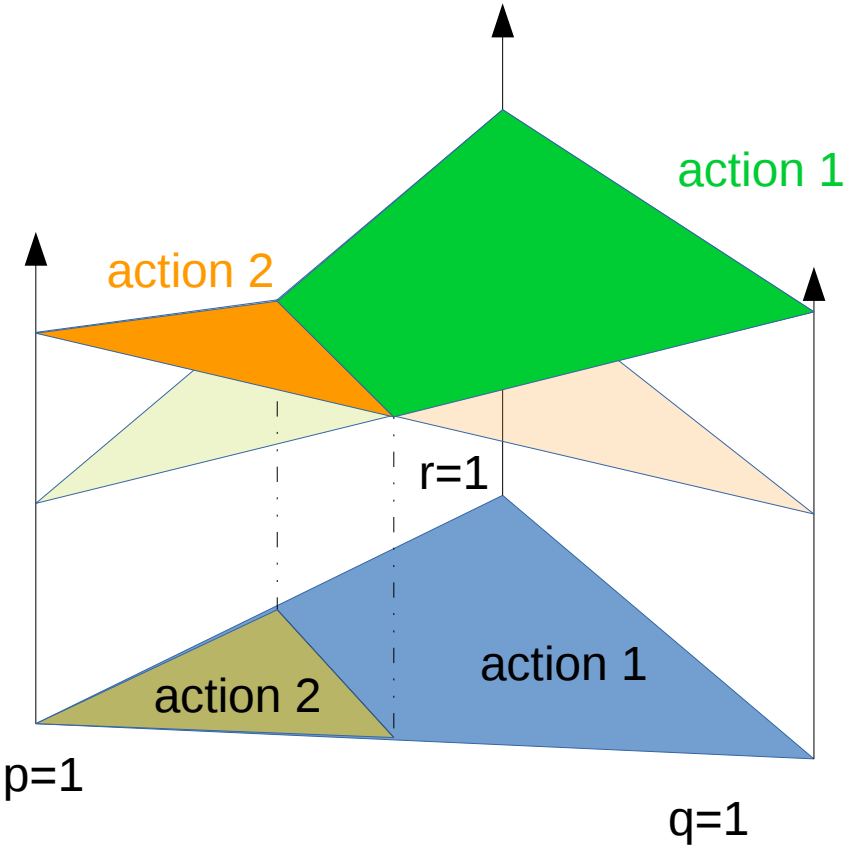


Belief space for 3 states

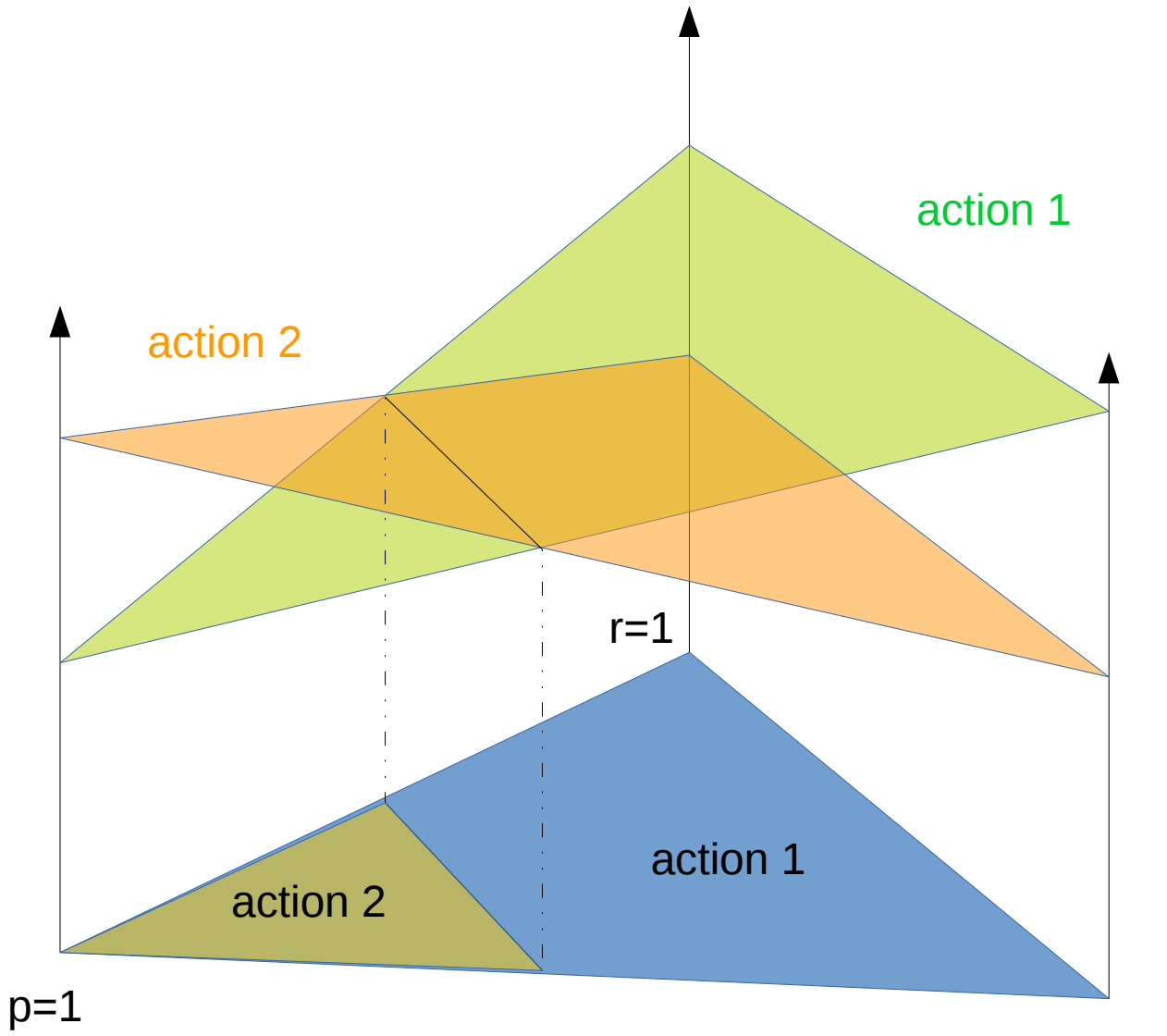
Often only the upper envelope is considered



2 states



The effective value function is piece-wise linear w.r.t. to the belief



Belief space for 3 states

How do we move in belief space?

It's the observations.

Different states have different likelihoods for observations.

Current belief is the prior

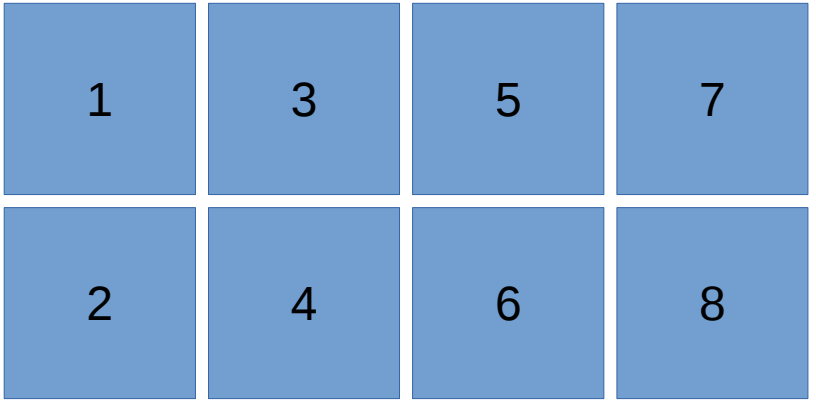
Bayes' formula yields a posterior, which serves as a new belief and as a new prior. (formulas later)

Particle filters can perform such operations in an efficient approximation, see also next slide

How do we move in belief space?

It's not only the observations. Actions affect real space, but have also an effect on the belief:

E.g. at t=1 p(1)=0.5 and p(2)=0.5



Moving two steps to the right may lead to p(5)=0.3 and p(6)=0.3 and p(3)=p(4)=p(7)=p(8)=0.1

An observation that is specific for 5 and 6 may concentrate the belief to these two states, see prev. slide

What happens in POMDPs?

- Agent moves in the environment by actions determined by the maximum over the value functions determined at the current belief state
- Value functions are updated by distributing the reward to states according to the belief
- New belief is determined by the effects of actions and of observations

A few more comments on POMDPs?

- Observations may come with an action or may be an action (that does not move the agent)
- For a time horizon of more than one step things get more complicated: One (linear) value function for each action-observation sequence
- We may neglect value functions that are dominated by other actions for any belief
- We may concentrate on points in belief space that actually occur from a given starting state
- We could ignore future uncertainty and assume certainty from the next step (similar to Q-learning)