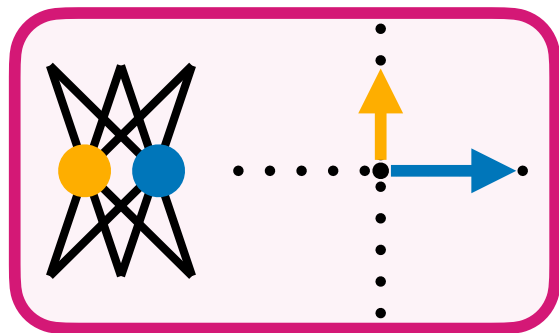
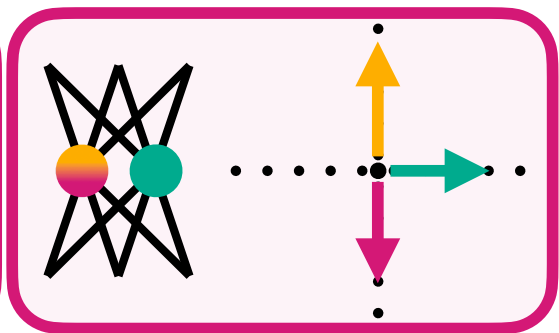


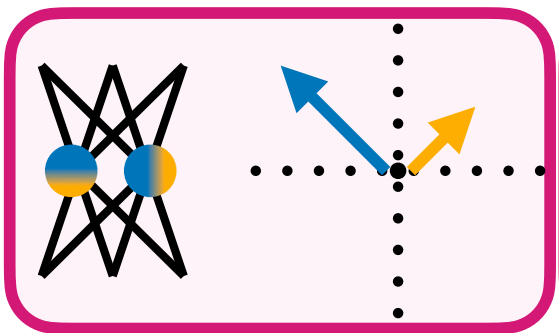
Privileged



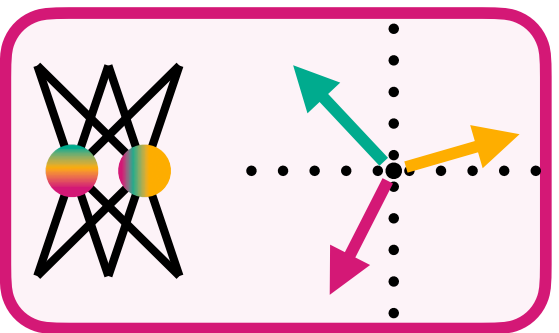
Privileged



Non-Privileged



Non-privileged



monosemantic neurons: ● ● ●

polysemantic neurons: ● ● ● ● ●