Change in the value function due to learning after taking action a^*:

$$\begin{aligned}
v(ba^*) - v(b) =& \sum_{b'} p(b' \mid b, a^*) \sum_{a} \pi(a \mid b') q(b', a) - \sum_{a} \pi(a \mid b) q(b, a) && (1)\\
=& \sum_{a} \big(\sum_{b'} p(b' \mid b, a^*) \pi(a \mid b') q(b', a) - \pi(a \mid b) q(b, a) \big)\\
=& \sum_{a} \big(\sum_{b'} p(b' \mid b, a^*) \pi(a \mid b') q(b', a) - \pi(a \mid b) q(b, a)\\
&+ \pi(a \mid b) \sum_{b'} p(b' \mid b, a^*) q(b', a) - \pi(a \mid b) \sum_{b'} p(b' \mid b, a^*) q(b', a) \big)\\
=& \sum_{a} \big(\sum_{b'} p(b' \mid b, a^*) \big(\pi(a \mid b') - \pi(a \mid b) \big) q(b', a)\\
&+ \pi(a \mid b) \sum_{b'} p(b' \mid b, a^*) \big(q(b', a) - q(b, a) \big) \big)\\
=& \sum_{b'} p(b' \mid b, a^*) \sum_{a} \big(\big(\pi(a \mid b') - \pi(a \mid b) \big) q(b', a)\\
&+ \pi(a \mid b) \big(q(b', a) - q(b, a) \big) \big)\\
& && (2)
\end{aligned}$$

$$\begin{aligned}
q(b', a) - q(b, a) =& \sum_{b''} p(b'' \mid b', a) \big[r(b', a) + \gamma v(b'') \big] && (3)\\
&- \sum_{b'} p(g' \mid b, a) \big[r(b, a) + \gamma v(b') \big]\\
=& r(b', a) + \gamma \sum_{b''} p(b'' \mid b', a) v(b'')\\
&- r(b, a) + \gamma \sum_{b'} p(b' \mid b, a) v(b')\\
=& \quad \underbrace{r(b', a) - r(b, a)}_{\text{Difference in the expected immediate return}} \quad \underbrace{+ \gamma \big[\sum_{b''} p(b'' \mid b', a) v(b'') - \sum_{b'} p(b' \mid b, a) v(b') \big]}_{\text{Difference in the expected future return}}
\end{aligned}$$