Change in the value function due to learning about a particular b':

$$
\begin{aligned}
v(b') - v(b) &= \sum_a \pi(a \mid b') q(b', a) - \sum_{a'} \pi(a \mid b) q(b, a) \\
&= \sum_a \left[(\pi(a \mid b') - \pi(a \mid b)) q(b', a) + \pi(a \mid b)(q(b', a) - q(b, a)) \right]
\end{aligned} \tag{1}
$$

$$
\begin{aligned}
q(b', a) - q(b, a) =& \sum_{b''} p(b'' \mid b', a) \big[r(b', a) + \gamma v(b'') \big] \\
& - \sum_{b'} p(g' \mid b, a) \big[r(b, a) + \gamma v(b') \big] \\
=& r(b', a) + \gamma \sum_{b''} p(b'' \mid b', a) v(b'') \\
& - r(b, a) + \gamma \sum_{b'} p(b' \mid b, a) v(b') \\
=& \quad \underbrace{r(b', a) - r(b, a)}_{\substack{\text{Difference in the expected} \\ \text{immediate return}}} \quad \underbrace{+ \gamma \big[\sum_{b''} p(b'' \mid b', a) v(b'') - \sum_{b'} p(b' \mid b, a) v(b') \big]}_{\substack{\text{Difference in the expected} \\ \text{future return}}}
\end{aligned} \tag{2}
$$

Note that we can write

$$
\begin{aligned}
cv(b'') - dv(b') =& cv(b'') - cv(b') + cv(b') - dv(b') \\
=& c(v(b'') - v(b')) + v(b')(c - d)
\end{aligned} \tag{3}
$$

Therefore

$$
\begin{aligned}
q(b', a) - q(b, a) =& r(b', a) - r(b, a) + \gamma \big[\sum_{b''} p(b'' \mid b', a) v(b'') - \sum_{b'} p(b' \mid b, a) v(b') \big] \\
=& r(b', a) - r(b, a) + \gamma \sum_{b''} p(b'' \mid b', a) \big[v(b'') - v(b') \big] \\
& + \gamma v(b') \big[\sum_{b''} p(b'' \mid b', a) - \sum_{b'} p(b' \mid b, a) \big]
\end{aligned} \tag{4}
$$

Note that the last term goes to zero. Therefore, substituting in:

$$
\begin{aligned}
v(b') - v(b) &= \sum_a \pi(a \mid b') q(b', a) - \sum_{a'} \pi(a \mid b) q(b, a) \\
&= \sum_a \big[\pi(a \mid b') - \pi(a \mid b) \big] q(b', a) \\
&+ \sum_a \pi(a \mid b) \big[r(b', a) - r(b, a) \big] \\
&+ \gamma \sum_a \pi(a \mid b) \big[\sum_{b''} p(b'' \mid b', a)(v(b'') - v(b')) \big]
\end{aligned}
$$

Unrolling:

$$
\begin{aligned}
v(b') - v(b) = &\sum_{i=0}^{\infty} \sum_{b' \in \mathcal{B}} \gamma^i P(b \to b', i, \pi(b)) \times \\
&\sum_a \Bigg(\underbrace{\big[\pi(a \mid b') - \pi(a \mid b)\big] q(b', a)}_{\text{Localised Gain}} + \underbrace{\mathbb{E}_{\pi(b)} \big[r(b', a) - r(b, a)\big]}_{\substack{\text{Accumuates into} \\ \text{long-term consequences}}} \Bigg)
\end{aligned} \tag{5}
$$

Equation 5 thus shows the non-local effect of policy change at a single belief state. Even though Gain is only non-zero at the exact (belief) location where we perform an update, this update nonetheless has long-lasting (discounted) consequences.