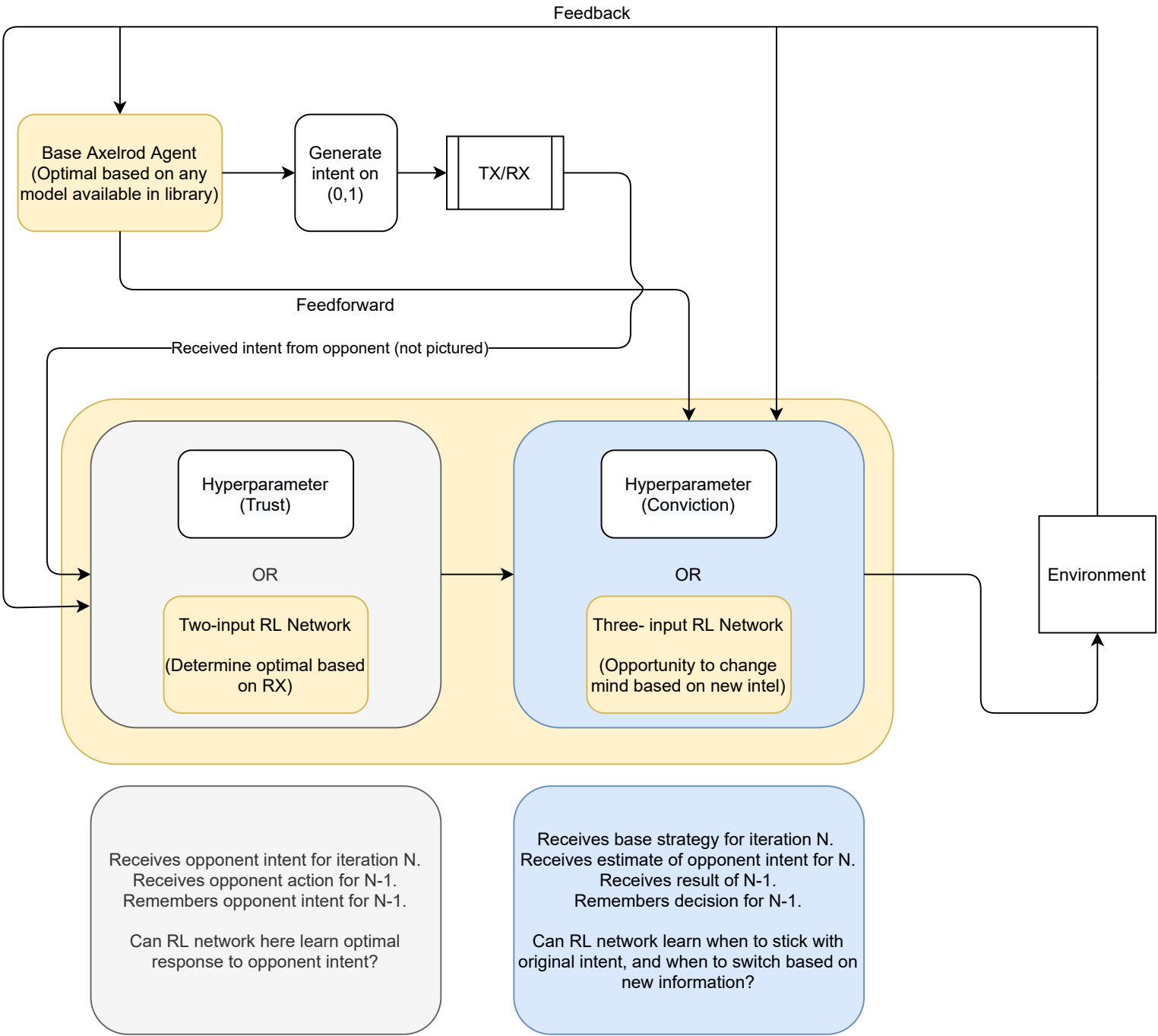
Feedback
Base Axelrod Agent (Optimal based on any model available in library)
Generate intent on (0,1)
TX/RX
Feedforward
Received intent from opponent (not pictured)
Hyperparameter (Trust)
OR
Two-input RL Network
(Determine optimal based on RX)
Hyperparameter (Conviction)
OR
Three- input RL Network
(Opportunity to change mind based on new intel)
Environment
Receives opponent intent for iteration N.
Receives opponent action for N-1.
Remembers opponent intent for N-1.
Can RL network here learn optimal response to opponent intent?
Receives base strategy for iteration N.
Receives estimate of opponent intent for N.
Receives result of N-1.
Remembers decision for N-1.
Can RL network learn when to stick with original intent, and when to switch based on new information?