

Theorem 2 Under free communication, consider a team of agents using a coordination policy: $\pi_{i\Sigma}(b_{i\bullet\Sigma}^t) \equiv \Omega_i^t$. If the domain-level policy π_A maximizes $V^T(\pi_A, \pi_\Sigma)$, then this combined policy is dominant over any other policies. In other words, for all policies, π'_A and π'_Σ , $V^T(\pi_A, \pi_\Sigma) \geq V^T(\pi'_A, \pi'_\Sigma)$.

Proof: Suppose we have some other coordination policy, π'_Σ , that specifies something other than complete communication (e.g., keeping quiet, lying). Suppose that there is some domain-level policy, π'_A , that allows the team to attain some expected reward, K , when used in combination with π'_Σ . Then, we can construct a domain-level policy, π_A , such that the team attains the same expected reward, K , when used in conjunction with the complete communication policy, π_Σ , as defined in the statement of Theorem 2.

The coordination policy, π'_Σ , produces a different set of belief states (denoted $b_{i\bullet\Sigma}^t$ and $b'_{i\Sigma\bullet}$) than those for π_Σ (denoted $b_{i\bullet\Sigma}^t$ and $b^t_{i\Sigma\bullet}$). In particular, we use state estimator functions, $SE'_{i\bullet\Sigma}$ and $SE'_{i\Sigma\bullet}$, as defined in Equations 2 and 3, to generate $b_{i\bullet\Sigma}^t$ and $b^t_{i\Sigma\bullet}$. Each belief state is a complete history of observation and communication pairs for each agent. On the other hand, under the complete communication of π_Σ , the post-communication state estimator function reduces to:

$$SE_{i\Sigma\bullet}(\langle \Omega^0, \dots, \Omega^{t-1}, \Omega_i^t \rangle, \Sigma^t) = \langle \Omega^0, \dots, \Omega^{t-1}, \Sigma^t \rangle$$

Since each agent's message is exactly its observation,

$$= \langle \Omega^0, \dots, \Omega^{t-1}, \Omega^t \rangle \quad (1)$$

Thus, π_A is defined over a different set of belief states than π'_A . In order to determine an equivalent π_A , we must first define a recursive mapping, m , that translates the belief states defined by π_Σ into those defined by π'_Σ :

$$m_i(b_{i\Sigma\bullet}^t)$$

The belief state at time t is a sequence of observations, which we can divide into the observations before time t and the observation at time t . The observations before time t correspond exactly to the belief state at time $t - 1$.

$$= m_i(\langle b_{i\Sigma\bullet}^{t-1}, \Omega^t \rangle)$$

The combined observation at time t includes agent i 's observation, as well as everyone else's observations.

$$= m_i(\langle b_{i\Sigma\bullet}^{t-1}, \langle \Omega_i^t, \Omega^t \rangle \rangle)$$

We can "distribute" the mapping function over the two components in the tuple. Under π'_Σ , the agents would not communicate their observations, but instead some other set of messages.

$$= \langle m_i(b_{i\Sigma\bullet}^{t-1}), \langle \Omega_i^t, \Sigma^t \rangle \rangle$$

We can break these messages down across the individual agents.

$$= \left\langle m_i(b_{i\Sigma\bullet}^{t-1}), \left\langle \Omega_i^t, \prod_{j \in \alpha} \Sigma_j^t \right\rangle \right\rangle$$

Each agent, j , selects its message based on following the communication policy, $\pi'_{j\Sigma}$, from its pre-communication belief state.

$$= \left\langle m_i(b_{i\Sigma\bullet}^{t-1}), \left\langle \Omega_i^t, \prod_{j \in \alpha} \pi'_{j\Sigma}(b^t_{j\bullet\Sigma}) \right\rangle \right\rangle$$

We can generate each agent's pre-communication belief state by applying the pre-communication state-estimator function to each agent's *previous* post-communication belief state and its most recent observation (which we know from its previous message, under full communication).

$$= \left\langle m_i(b_{i\Sigma\bullet}^{t-1}), \left\langle \Omega_i^t, \prod_{j \in \alpha} \pi'_{j\Sigma}(SE'_{j\bullet\Sigma}(b_{j\Sigma\bullet}^{t-1}, \Omega_j^t)) \right\rangle \right\rangle$$

Finally, we can compute each agent's previous post-communication belief state by again applying the mapping function.

$$= \left\langle m_i(b_{i\Sigma\bullet}^{t-1}), \left\langle \Omega_i^t, \prod_{j \in \alpha} \pi'_{j\Sigma}(SE'_{j\bullet\Sigma}(m_j(b_{j\Sigma\bullet}^{t-1}), \Omega_j^t)) \right\rangle \right\rangle \quad (2)$$

Given this mapping, we then specify: $\pi_{iA}(b_{i\Sigma\bullet}^t) = \pi'_{iA}(m_i(b_{i\Sigma\bullet}^t))$. Executing this domain-level policy, in conjunction with the coordination policy, π_Σ , results in the identical behavior as execution of the alternate policies, $\pi_{A'}$ and $\pi_{\Sigma'}$. Therefore, the team following the policies, π_A and π_Σ will achieve the same expected value of K , as under $\pi_{A'}$ and $\pi_{\Sigma'}$. \square