Supplementary Materials

This supplemental document provides the following additional contents to support the main paper:

- A Overview of notation used in the paper
- B Additional quantitative evaluation
- C Interpretation of $m^1_{O\rightarrow N}$ for 1-ON
- D Interpretation of `S-Comm` using vocabulary size 3
- E Interpretation of `U-Comm` and `S-Comm` for 2-ON
- G Additional analyses to check for information content of messages
- F Episode map visualizations
- H Implementation details

A. Notation overview

Table 1 provides a summary of the definitions of important notations used in the paper and this supplementary document.

B. Additional quantitative evaluation

We report only the PPL and PROGRESS metrics in the main paper. Table 2 summarizes the complete set of evaluation metrics we use: PROGRESS, PPL, SUCCESS and SPL. The trends in SUCCESS and SPL are similar to those in PROGRESS and PPL. We also perform generalization experiments by evaluating models trained on 3-ON on the more difficult 4-ON and 5-ON tasks. Table 2 summarizes those results as well. We observe that `S-Comm` outperforms `U-Comm` in these generalization experiments as well.

Table 3 shows the effect of increasing message length on the task performance. 2-dimensional message results are repeated from **??**. `S-Comm` still outperforms `U-Comm` on equal message lengths. Overall, there are small improvements with increasing message dimension. We hypothesize that A_O can encode the goal location in 2-dimensional messages, thus higher-dimensional messages provide small improvements.

C. Interpretation of $m^1_{O\rightarrow N}$ for 1-ON

In the CoMON task, A_N has knowledge of the goal to visit while A_O has knowledge of the semantic map (with goal positions) as well as the position and orientation of A_O. In the main paper, we showed that $m^2_{O\rightarrow N}$ is used to communicate the location of the goal. Here we consider what information is in the initial message sent from A_O to A_N ($m^1_{O\rightarrow N}$) for 1-ON. Note that in the case of 1-ON, there is only one goal, so A_O can already send information about that goal without waiting for $m^1_{N\rightarrow O}$. Using similar methods as we employed in the main paper for $m^2_{O\rightarrow N}$, we find that $m^1_{O\rightarrow N}$ is also used to communicate the location of the goal relative to A_N for `U-Comm` and `S-Comm`.

C.1. Interpretation of $m^1_{O\rightarrow N}$ in `U-Comm`

Figure 1 shows the distribution of $m^1_{O\rightarrow N}$ w.r.t. the relative coordinates of the goal object from A_N, using a similar visualization as for $m^2_{O\rightarrow N}$ (Figure 5 in the main paper). We note there is a correlation between $m^1_{O\rightarrow N}$ and the location of the current goal object: with the first element indicating whether the goal is to the left of the agent, and the second element whether the goal is to the right. To quantify the relation, we again fit linear probes. As the target of the linear probe, we bin the angles into 8 bins each of 45° (see dashed lines in Figure 1). Our probe attains classification accuracy of 51% (compared to chance accuracy of 12.5%) supporting our hypothesis that $m^1_{O\rightarrow N}$ includes information about the location of the current goal object relative to A_N.

C.2. Interpretation of $m^1_{O\rightarrow N}$ in `S-Comm`

As in the main paper (Section 6.2), we perform a similar analysis for $m^1_{O\rightarrow N}$ as for $m^2_{O\rightarrow N}$, where we use thresholds to group the messages into Δ_1, Δ_2, and Δ_3. Figure 2 plots the distribution of each symbol w.r.t. the relative location of the current goal relative to A_N (similar to Figure 6 in the main paper). We observe that $m^1_{O\rightarrow N}$ is again used to convey the goal object location, but the correlation between the communicated message and the goal object location is weaker than that of $m^2_{O\rightarrow N}$. This is evident from the higher overlap of the regions corresponding to each symbol (compared to Figure 6 in the main paper). This observation is confirmed by the lower classification accuracy of 83% (vs 89% for $m^2_{O\rightarrow N}$) after training a random forest classifier to predict the communicated symbol from the (x, y) coordinate of the current goal object.

Notation	Description	Notation	Description
m-ON	Episode with m ordered object goals	$\hat{b}_N$	Initial belief of A_N
G	Sequence of goal objects	b_O	Final belief of A_O
A_O	Oracle agent	b_N	Final belief of A_N
A_N	Navigator agent	v_a	Embedding of previous action a_{t-1}
M	Oracle map in global frame	s	Final state representation
o_t	Egocentric RGBD frames	$m^r_{N\to O}$	Message sent by A_O to A_N in round r
g_t	Current goal object one-hot vector	$m^r_{O\to N}$	Message sent by A_N to A_O in round r
a_t	Action taken by the agent	r_t	Reward at time-step t
`U-Comm`	Unstructured communication	r^{goal}	Reward for finding a goal
`S-Comm`	Structured communication	r^{closer}	Reward for moving closer to goal
E	Oracle map in egocentric frame	$r^{\text{time penalty}}$	Time penalty reward
v_o	RGBD features	w_i	Embedding of word i
v_g	Embedding of one-hot goal vector g_t	p_i	Probability for word i
$\hat{b}_O$	Initial belief of A_O	Δ_i	Communication symbol i

Table 1. **Summary of notation.** Subscript t denotes the corresponding notation at time step t

	Progress (%)					PPL (%)					Success (%)					SPL (%)				
	1-ON	2-ON	3-ON	4-ON	5-ON	1-ON	2-ON	3-ON	4-ON	5-ON	1-ON	2-ON	3-ON	4-ON	5-ON	1-ON	2-ON	3-ON	4-ON	5-ON
`NoCom`	56	39	26	10	7	35	26	16	7	5	56	30	10	7	2	35	18	5	3	1
`Rand U-Comm`	59	40	28	7	5	36	28	18	3	2	58	33	12	0	0	32	20	6	0	0
`Rand S-Comm`	50	31	24	16	10	33	24	16	11	6	50	30	9	6	1	33	16	5	3	1
`U-Comm`	**87**	77	63	41	26	60	51	39	23	13	**87**	57	53	23	7	60	43	40	13	3
`S-Comm`	85	**80**	**70**	**50**	**35**	**67**	**59**	**50**	**32**	**22**	85	**65**	**58**	**32**	**14**	**67**	**46**	**45**	**20**	**9**
`OracleMap`	89	80	70	45	26	74	64	52	28	14	89	69	61	27	8	74	49	42	16	4

Table 2. **Additional quantitative metrics on** 1-ON, 2-ON **and** 3-ON **tasks and generalization to** 4-**ON and** 5-**ON.** All agents are trained on 3-ON and evaluated on the task indicated in each column.

	Dim	Progress (%)			PPL (%)		
		1-ON	2-ON	3-ON	1-ON	2-ON	3-ON
`U-Comm`	2	87	77	63	60	51	39
	3	88	78	67	66	55	45
	4	88	79	68	66	57	46
`S-Comm`	2	85	**80**	**70**	67	59	50
	3	89	78	**70**	**72**	57	52
	4	**90**	**80**	**70**	67	**60**	**54**

Table 3. **Effect of message length on performance.** `S-Comm` outperforms `U-Comm` on similar message length and there is a slight improvement with increasing message dimension.

D. Interpretation of `S-Comm` for vocabulary size 3

We provide details of the analysis of `S-Comm` with vocabulary of size 3. Similar to our analysis for vocabulary size 2 (see section 6.2 in main paper), we bin the messages probabilities based on the observed probabilities. Due to the larger vocabulary size, we bin the messages into six classes (vs 3 classes for vocabulary size of 2): Δ_1, Δ_2, Δ_3, Δ_4, Δ_5 or Δ_6. See H.4 for more details about the binning process. When we examine the messages, we see a consistent pattern as we observed for vocabulary size 2.

What does A_N tell A_O in $m^1_{N\to O}$? Here also, we observe that A_N uses $m^1_{N\to O}$ to convey the goal object to A_O. A_N send Δ_1 when the goal object is a red, green, pink, or cyan cylinder. It sends Δ_2 for blue and yellow cylinders, and it sends Δ_3 otherwise. We find that Δ_4, Δ_5, Δ_6 are not used for $m^1_{N\to O}$, and are only used in $m^2_{O\to N}$ and $m^1_{O\to N}$.

What does A_O tell A_N in $m^2_{O\to N}$? We perform the same interpretation analysis as we did for vocabulary size 2 in the main paper. We again observe that A_O utilizes $m^2_{O\to N}$ to convey the goal location to A_N (see Figure 3). Because of the availability of more communication symbols in vocabulary size 3, A_O send more fine-grained information about

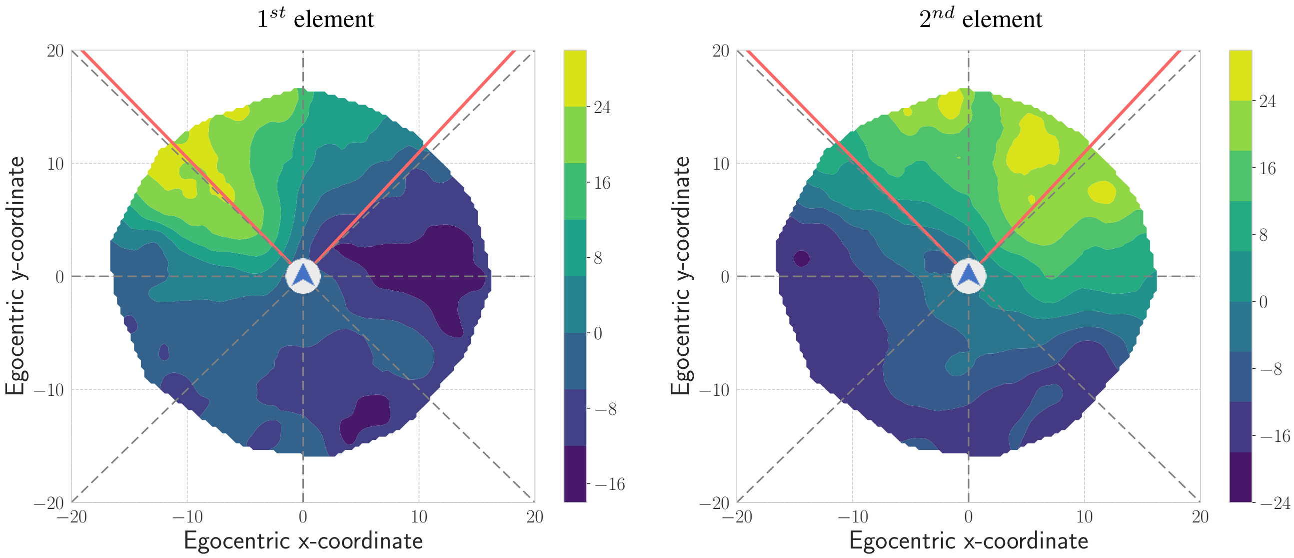

Figure 1. **Egocentric visualization of `U-Comm` communication symbol $m^1_{O\to N}$.** The two plots visualize the value of the first and second element of the message plotted w.r.t. the relative coordinates of the goal object from A_N. The navigator agent A_N is facing the +y axis and its field-of-view is marked with red lines. The plot on the left corresponds to the 1st dimension of the message, while the plot on the right corresponds to the 2nd dimension. The value of each dimension is indicated by the color hue.

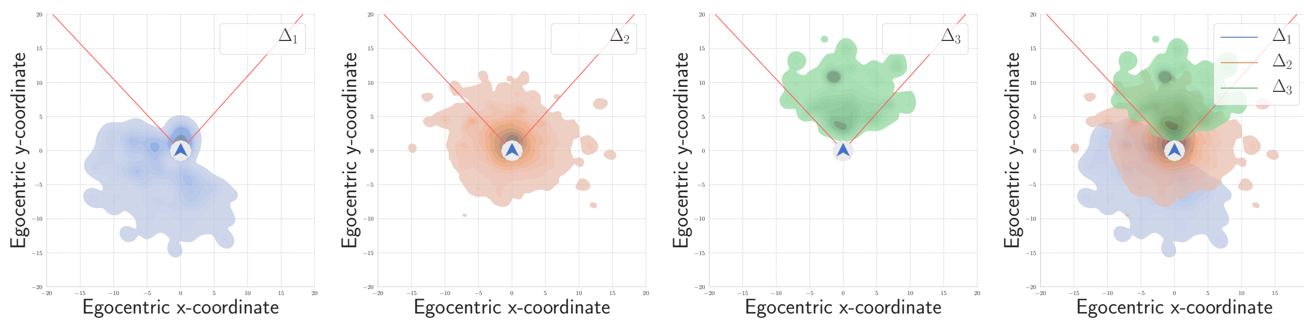

Figure 2. **Egocentric visualization of `S-Comm` communication symbol $m^1_{O\to N}$.** The plots show the relative coordinates of the current goal object from A_N's perspective when A_O communicates the symbol through `S-Comm` with vocabulary size two. The navigator agent (A_N) is facing the +y axis and its field-of-view is marked with red lines. Data points are accumulated across all validation episodes, and we plot contour lines of the bivariate density distribution. Each data point is a message with (x, y) coordinates determined from the coordinates of the current goal object in A_N's egocentric reference frame when the message was sent. The first three plots are for each communication symbol, and the right-most combines all symbols. Note how each symbol represents distinct regions that are egocentrically organized around the agent: Δ_1 captures 'behind and not visible', Δ_2 corresponds mostly to 'close, in front', and Δ_3 is 'farther in front'.

the regions. Similar to vocabulary size 2 (main paper section 6.2), we observe that more symbols are allocated to the front of the agent than at its back.

What does A_O tell A_N in $m^1_{O\to N}$? For this message, our observations are again consistent with those of vocabulary size 2 (see Figure 4). A_O sends different symbols for different goal locations, but there is more overlap between the regions allocated to the symbols as compared to that in $m^2_{O\to N}$.

E. Interpretation for 2-ON

Most of our analysis thus far has focused on 1-ON. Here we analyze what is communicated in `U-Comm` and `S-Comm` for 2-ON using the same methodology.

E.1. Interpretation of `U-Comm` for 2-ON

What does A_N tell A_O in $m^1_{N\to O}$? We observe that the distribution of $m^1_{N\to O}$ is similar to that in Figure 4 of the main paper. This is expected as A_N would send similar message irrespective of the number of goals in an episode.

What does A_O tell A_N in $m^2_{O\to N}$? In Figure 5, we show show the distribution of $m^2_{O\to N}$ against the current object goal in the spatial reference frame defined by the position and orientation of A_N (egocentric frame) at the environment step when the message was sent. Our observations are consistent with what we observed for 1-ON. $m^2_{O\to N}$ is used to convey the goal location to A_N.

E.2. Interpretation of `S-Comm` for 2-ON

In this setting we use a vocabulary size of 2 and group the messages into three symbols as in the main paper. Because the number of symbols is less than the number of possible goals, we observe that the agents use a partitioning scheme when sending messages. This phenomenon has been observed in Kottur et al. and is consistent with game theory results.

What does A_N tell A_O in $m^1_{N\to O}$? Here also, A_O sends similar $m^1_{N\to O}$ as in 1-ON. That is, A_O sends Δ_1 when the goal object is a red, white or black cylinder, and sends Δ_2 otherwise. A_N partitions the goal objects into two sets: P_1 with 3 categories and P_2 with 5 categories.

What does A_O tell A_N in $m^2_{O\to N}$? As $m^1_{N\to O}$ only sends Δ_1 or Δ_2, A_O cannot infer the precise current goal object. If both of these objects lie in P_2, A_N would send Δ_2 to A_O throughout the episode. Therefore, A_O would not know which of the 2 objects A_N is looking for at the moment. Instead, if one of the target objects lies in P_1 and the other in P_2, A_O can infer the current target object A_N is looking for. We plot the message $m^2_{O\to N}$ for the two cases separately. In Figure 6, first row represents the case when the current goal can be distinguished from $m^1_{N\to O}$. Note that the current goal is said to be *distinguishable* from $m^1_{N\to O}$ if the two goals for the episode lie in separate partitions P_1 and P_2. $m^2_{O\to N}$ correlates more strongly with the location of the current goal in the former case, where A_O could infer the current goal before sending $m^2_{O\to N}$. This is reflected in the symbols being more well separated in the first row of Figure 6 than in the second row. This can be observed by the overlaps between symbols. Distinguishable goals have less overlap between symbol regions as compared to indistinguishable goals. To quantify the separation of symbols in the two plots, we also train a random forest classifier to predict the communication symbol given the x,y coordinates of the symbol in the plots as input. The prediction accuracy for distinguishable goals is 84% and for indistinguishable goals it is 76%.

F. Episode map visualizations

In Figure 7, we provide a visualization of egocentric observations and map state for `S-Comm` at several points on the trajectory to show correlations between the communication symbols for $m_{N\to O}$ (shown on the trajectory) and what the agent observes at each position. Similarly in Figure 8, we show the correlations for $m_{O\to N}$.

G. Are messages conveying other information?

We investigated other information that the messages might be conveying, but did not find a strong signal. We checked if $m^1_{O\to N}$ or $m^2_{O\to N}$ conveys the optimal action and if $m^1_{N\to O}$ conveys whether the current goal is in A_N's view. We also checked whether messages from A_N to A_O contain direction, and messages from A_O to A_N contain color, and did not find any correlations.

H. Implementation details

H.1. Architecture details

Here we report the architectural details. A_O has an oracle map M of spatial dimension 300×300. This contains occupancy and goal object information. M is converted to egocentric map E of spatial dimension 45×45. Each of the occupancy and goal object information is converted to 16 dimensional embeddings for each grid location so the map is of dimension $45\times 45\times 32$. This is passed through a map encoder comprising of a two layered CNN and a linear layer to obtain $256-$dimensional belief $\hat{b}_O$. b_O is a $256-$dimensional vector as well.

RGBD observations of A_O are passed through an image encoder. It consists of three CNN layers and a linear layer to obtain an image embedding v_o of shape 512. The current goal embedding v_g and previous action embedding v_a are both $16-$dimensional vectors. The belief $\hat{b}_N$ and b_O are of shape 512. The state representation vector s is of shape 528.

H.2. Details about random baselines

Here, we present the implementation details for `Rand U-Comm` and `Rand S-Comm`. In `Rand U-Comm`, we replace the message by a random vector sampled from a multivariate gaussian distribution with mean and variance equal to the mean and variance of the corresponding message sent in the validation set. For `Rand S-Comm`, we replace the message by random probabilities sampled from a random multinomial probability vectors and these probabilities sum up to 1.

H.3. `S-Comm` classifier implementation

To establish the existence of various correlations between the communication symbols exchanged between the agents in `S-Comm`, we train random forest classifiers that predict the communication symbol given a quantity Q as input. We report the classification accuracy for $m^1_{N\to O}$ and $m^2_{O\to N}$ in Section 6.2 and for $m^1_{O\to N}$ above. For all of these, the data for training/evaluating the classifier is obtained by evaluating the model on the validation set of 1,000 episodes and accumulating the relevant metrics at each environment step across the 1,000 episodes. At each environment step, we log the following: $\{m^1_{N\to O}, m^1_{O\to N}, m^2_{O\to N}$, current object goal category, relative location of current goal in A_N's egocentric reference frame.} We first balance the dataset such that each symbol Δ_i has equal number of training examples. The collected data, where each data point corresponds to an environment step, is divided into train and val sets in 3/1 ratio. The classifier is then trained to predict the communication symbol Δ_i from quantity q using the train set. We report the classification accuracy on the val set.

H.4. Binning of probabilities in `S-Comm`

Here, we describe the implementation of binning to create the discrete symbols used in our interpretation of `S-Comm`.

Vocabulary size 2. Let the probability vector output by the final softmax layer of communication module be $\boldsymbol{p} = [p_1, p_2]$ and let the binned vector be $\boldsymbol{d}$. If $p_1 < 0.2$, $\boldsymbol{d} = [0, 1]$; if $p_1 > 0.8$, $\boldsymbol{d} = [1, 0]$; and if $0.2 \leq p_1 \leq 0.8$, $\boldsymbol{d} = [0.5, 0.5]$. As such, each agent sends one of the three categorical vectors during each round of communication.

Vocabulary size 3. Here, the model outputs a probability vector $\boldsymbol{p}$ of length 3: $[p_1, p_2, p_3]$. The procedure for obtaining the binned vector $\boldsymbol{d}$ is described below:

$$\boldsymbol{d} = \begin{cases} [1,0,0] & \text{if } p_1 > 0.75 \\ [0,1,0] & \text{if } p_2 > 0.75 \\ [0,0,1] & \text{if } p_3 > 0.75 \\ [0,0.5,0.5] & \text{if } \max(p_1,p_2,p_3) < 0.75 \\ & \text{and } p_1 < p_2, p_3 \\ [0.5,0,0.5] & \text{if } \max(p_1,p_2,p_3) < 0.75 \\ & \text{and } p_2 < p_1, p_3 \\ [0.5,0.5,0] & \text{if } \max(p_1,p_2,p_3) < 0.75 \\ & \text{and } p_3 < p_2, p_1 \end{cases}$$

Under this formulation, each agent can be considered to send only a discrete symbol to the other agent during communication.

References

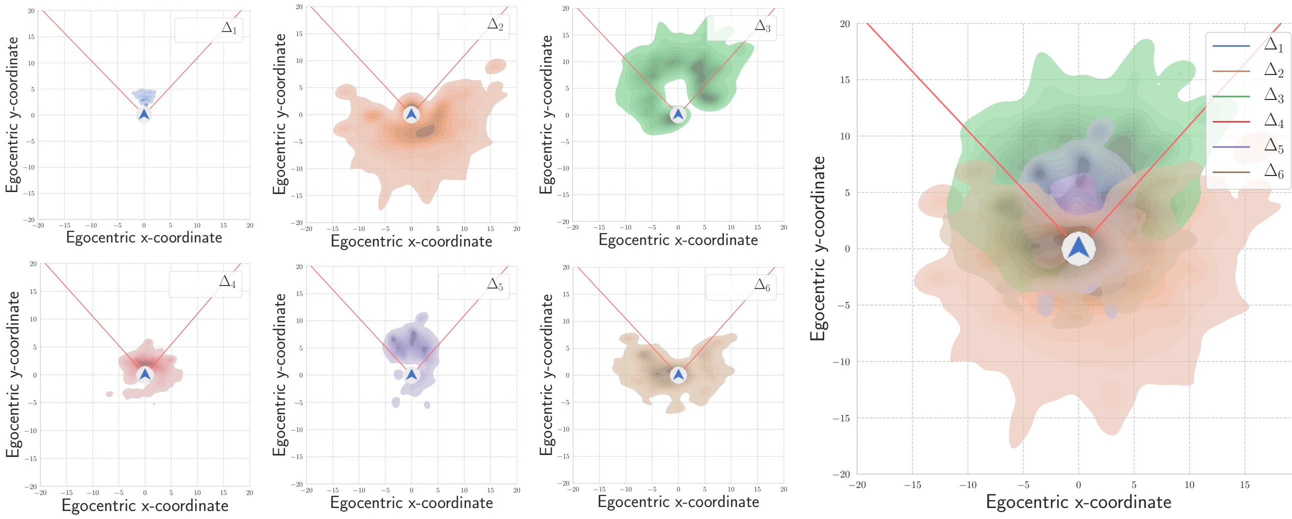

Figure 3. **Egocentric visualization of `S-Comm` communication symbol $m^2_{O\to N}$ for vocabulary size 3.** The plots show the relative coordinates of the current goal object from A_N's perspective when A_O communicates the symbol through `S-Comm` with vocabulary size three. The navigator agent (A_N) is facing the +y axis and its field-of-view is marked with red lines. Data points are accumulated across all validation episodes, and we plot contour lines of the bivariate density distribution. Each data point is a message with (x, y) coordinates determined from the coordinates of the current goal object in A_N's egocentric reference frame when the message was sent. The six plots on the left are for each communication symbol, and the right-most combines all symbols. Note how each symbol represents distinct regions that are egocentrically organized around the agent.

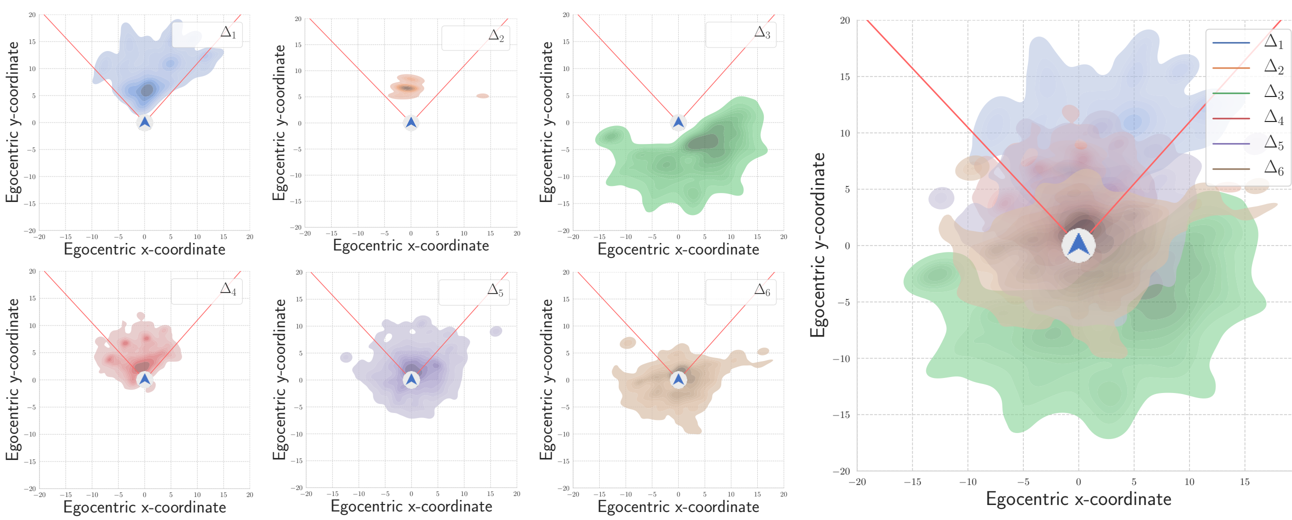

Figure 4. **Egocentric visualization of `S-Comm` communication symbol $m^1_{O\to N}$ for vocabulary size 3.** The plots show the relative coordinates of the current goal object from A_N's perspective when A_O communicates the symbol through `S-Comm` with vocabulary size two. The navigator agent (A_N) is facing the +y axis and its field-of-view is marked with red lines. Data points are accumulated across all validation episodes, and we plot contour lines of the bivariate density distribution. Each data point is a message with (x, y) coordinates determined from the coordinates of the current goal object in A_N's egocentric reference frame when the message was sent. The first three plots are for each communication symbol, and the right-most combines all symbols.

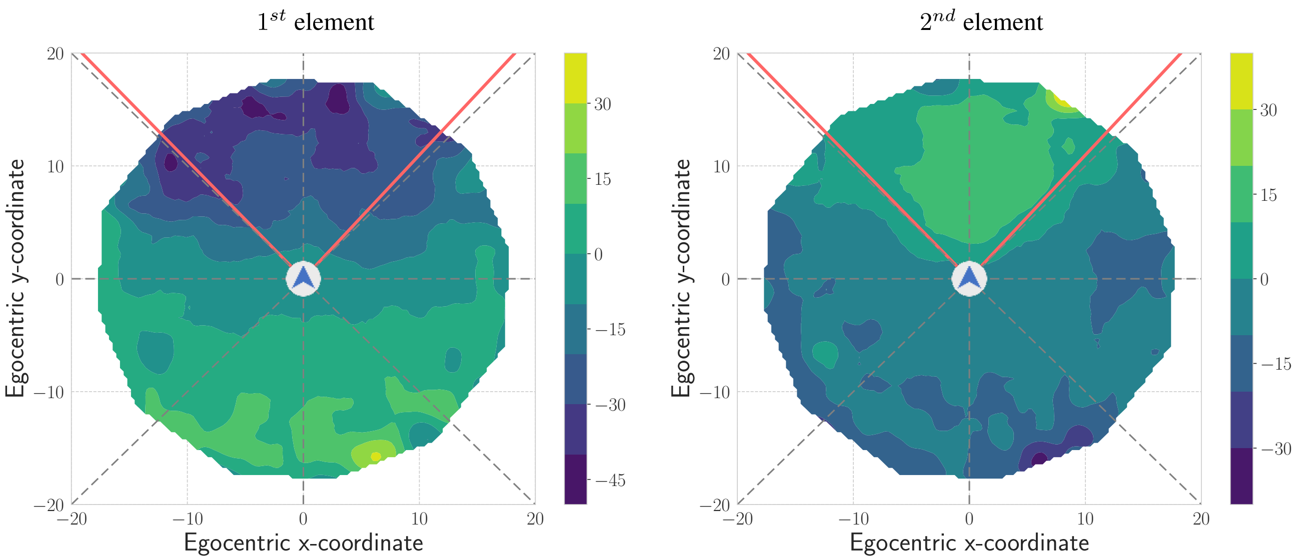

Figure 5. **Egocentric visualization of `U-Comm` communication symbol $m^2_{O\to N}$ for 2-ON.** The two plots visualize the value of the first and second element of the message plotted w.r.t. the relative coordinates of the goal object from A_N. The navigator agent A_N is facing the +y axis and its field-of-view is marked with red lines. The plot on the left corresponds to the 1st dimension of the message, while the plot on the right corresponds to the 2nd dimension. The value of each dimension is indicated by the color hue.

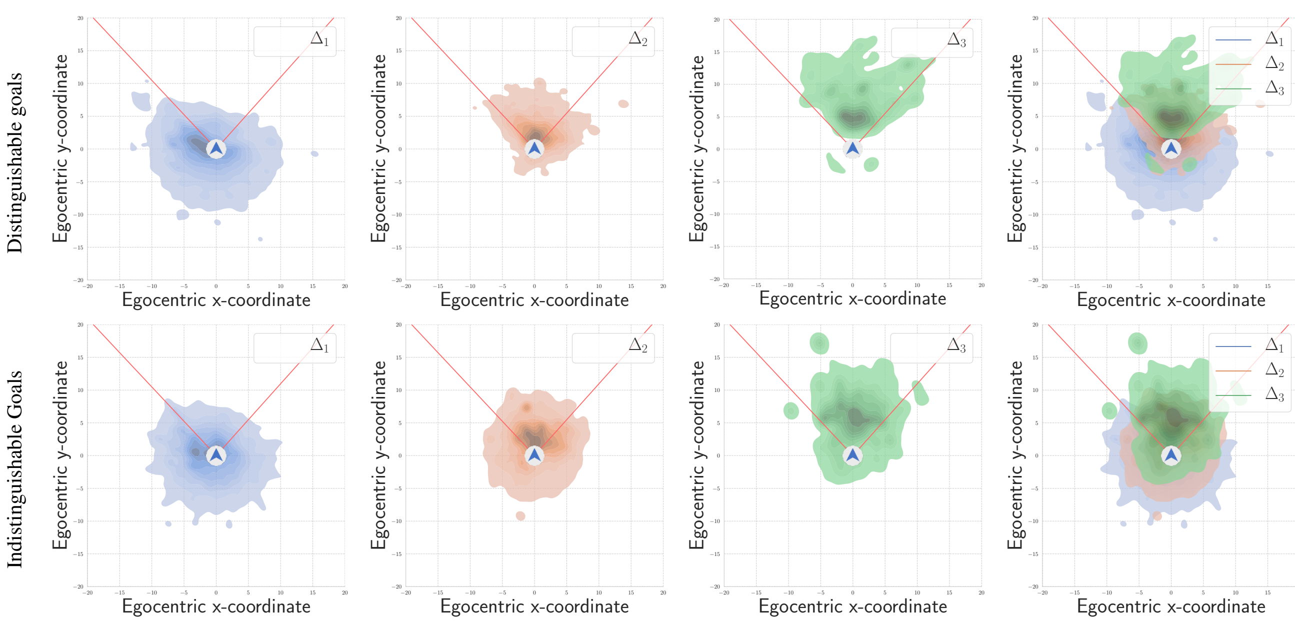

Figure 6. **Egocentric visualization of `S-Comm` communication symbol $m^2_{O\to N}$.** First and second row show the case when the goals are distinguishable and indistinguishable by A_O respectively. The plots show the relative coordinates of the current goal object from A_N's perspective when A_O communicates the symbol through `S-Comm` with vocabulary size two. The navigator agent (A_N) is facing the +y axis and its field-of-view is marked with red lines. Data points are accumulated across all validation episodes, and we plot contour lines of the bivariate density distribution. Each data point is a message with (x, y) coordinates determined from the coordinates of the current goal object in A_N's egocentric reference frame when the message was sent. The first three plots are for each communication symbol, and the right-most combines all symbols. Notice that first row symbols have lesser overlap than the second row symbols.

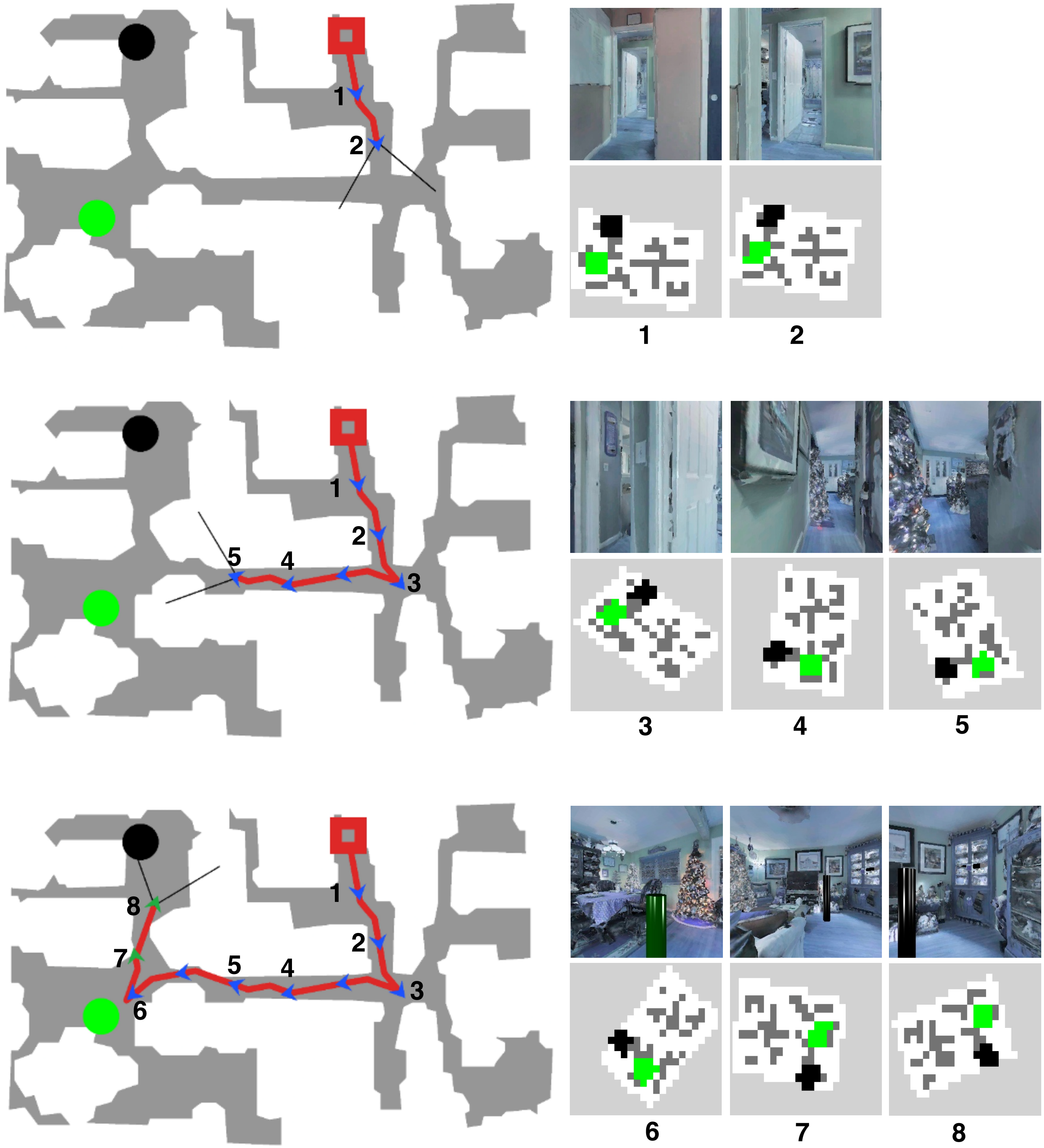

Figure 7. **Example navigation episode with communication message $m_{N \to O}$ on the agent trajectory for `S-Comm`.** The message $m_{N \to O}$ is depicted by the color of the arrow symbol at various points on A_N's trajectory on the top-down map. The sequence of maps from top-to-bottom visualizes the trajectory at different points in time. Egocentric observations and map representations at specific agent positions are given to the right of each map. Note the changed communication symbol (from blue agent symbol to green) after the first green goal is found and the agent proceeds to the next black goal.

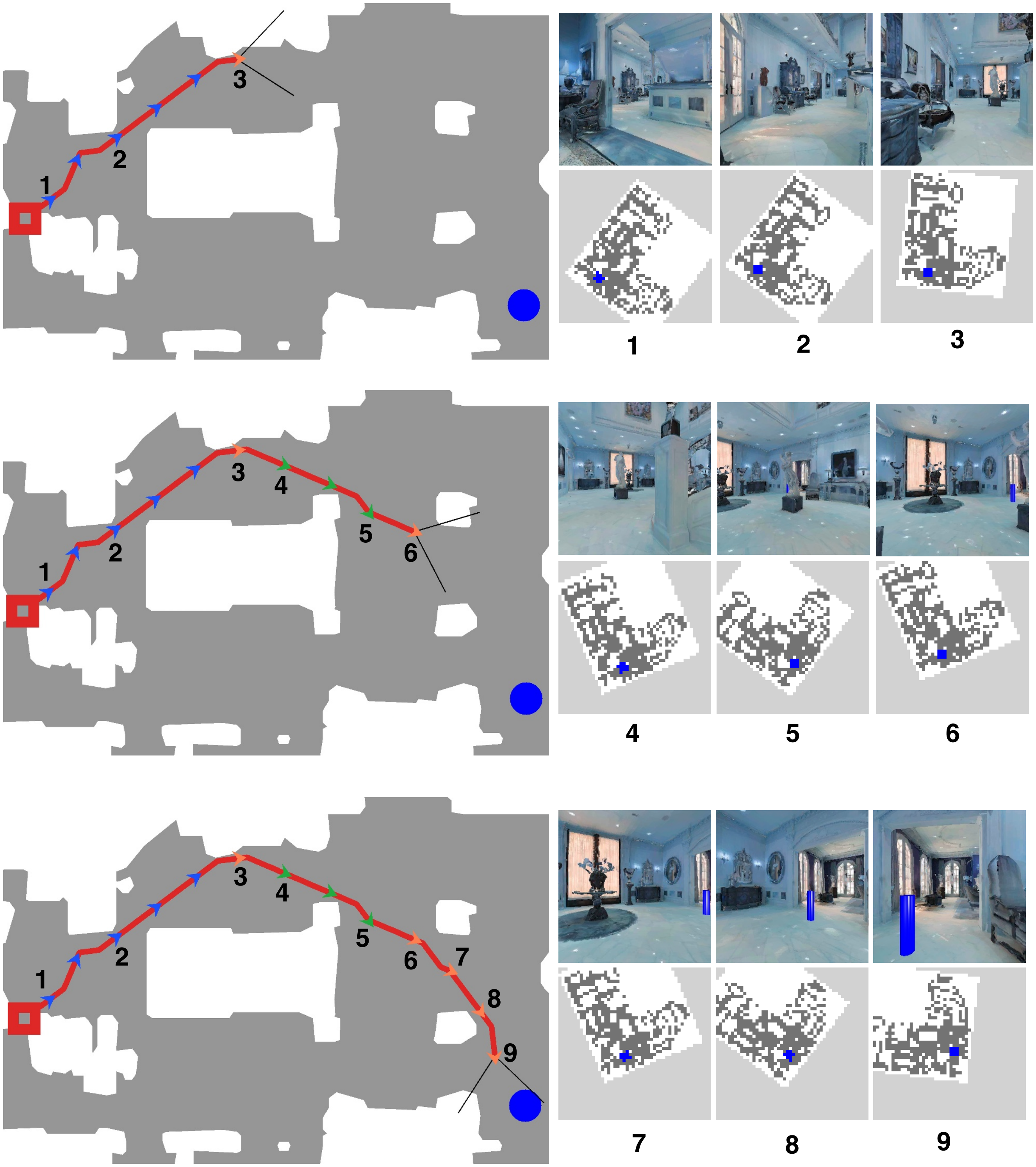

Figure 8. **Example navigation episode with communication message $m_{O\to N}$ on the agent trajectory for `S-Comm`.** The message $m_{O\to N}$ is depicted by the color of the arrow symbol at various points on A_N's trajectory on the top-down map. The sequence of maps from top-to-bottom visualizes the trajectory at different points in time. Egocentric observations and map representations at specific agent positions are given to the right of each map. Note how the communication symbol changes as the relative location of the goal object with respect to the agent changes: when the goal is not ahead of the agent, it is blue; when the goal is ahead of the agent but far away, it is green; and when the goal is in front of the agent, it is orange.